REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704-0188

1. AGENCY USE ONLY (Leave blank)	2. REPORT DATE	3. REPORT TYPE AND DATES COVERED
	May 93	final; 6 Jun 91 - 14 May 93

4. TITLE AND SUBTITLE
Tertiary structural studies of myotoxin a from Crotalus viridis viridis venom by nuclear magnetic resonance. (Masters thesis)

5. FUNDING NUMBERS

6. AUTHOR(S)
O'Keefe, Michael P.

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)
U.S. Army Student Detachment
w/ duty at Arizona State University
Tempe, AZ 85287

8. PERFORMING ORGANIZATION REPORT NUMBER

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)
U.S. Army Student Detachment
Ft. Benjamin Harrison, IN 46216

10. SPONSORING/MONITORING AGENCY REPORT NUMBER

11. SUPPLEMENTARY NOTES
Thesis prepared for M.S.

12a. DISTRIBUTION/AVAILABILITY STATEMENT
Available to public.

DISTRIBUTION STATEMENT A
Approved for public release;
Distribution Unlimited

12b. DISTRIBUTION CODE

13. ABSTRACT (Maximum 200 words)

Myotoxin a, a 42-residue protein from the venom of the prairie rattlesnake (Crotalus viridis viridis), has been studied in aqueous solution by proton nuclear magnetic resonance (NMR) spectroscopy. and a general tertiary structure has been determined. Myotoxin a is one of a family of highly homologous myotoxins that cause localized tissue myonecrosis upon envenomation and whose structures are highly constrained by three disulfide linkages. Eighty-six relevant distance constraints derived from nuclear Overhauser enhancement spectroscopy (NOESY) experiments were employed in distance geometry calculations. A superimposed subset of the best refined structures yielded a medium resolution (backbone atoms' root mean square distance of 2.5 Å) tertiary conformation. The structure consists of three strands of anti-parallel beta sheet bound by three disulfide bonds and connected by short loops and turns, including a modified type VI (cis-proline) turn.

14. SUBJECT TERMS	15. NUMBER OF PAGES
myotoxin; nuclear magnetic resonance; distance geometry; rattlesnake; venom	104
	16. PRICE CODE

17. SECURITY CLASSIFICATION OF REPORT	18. SECURITY CLASSIFICATION OF THIS PAGE	19. SECURITY CLASSIFICATION OF ABSTRACT	20. LIMITATION OF ABSTRACT
UNCLASSIFIED	UNCLASSIFIED	UNCLASSIFIED	UL

TERTIARY STRUCTURAL STUDIES OF MYOTOXIN *A* FROM *CROTALUS VIRIDIS VIRIDIS* VENOM BY NUCLEAR MAGNETIC RESONANCE

by

Michael P. O'Keefe

A Thesis Presented in Partial Fulfillment
of the Requirements for the Degree
Master of Science

ARIZONA STATE UNIVERSITY

May 1993

TERTIARY STRUCTURAL STUDIES OF MYOTOXIN *A*

FROM *CROTALUS VIRIDIS VIRIDIS* VENOM

BY NUCLEAR MAGNETIC RESONANCE

by

Michael P. O'Keefe

has been approved

April 1993

APPROVED:

Chairperson

Supervisory Committee

ACCEPTED:

Department Chairperson

Dean, Graduate College

ABSTRACT

Myotoxin *a*, a 42-residue protein from the venom of the prairie rattlesnake (*Crotalus viridis viridis*), has been studied in aqueous solution by proton nuclear magnetic resonance (NMR) spectroscopy, and a general tertiary structure has been determined. Myotoxin *a* is one of a family of highly homologous myotoxins that cause localized tissue myonecrosis upon envenomation and whose structures are highly constrained by three disulfide linkages. Eighty-six relevant distance constraints derived from nuclear Overhauser enhancement spectroscopy (NOESY) experiments were employed in distance geometry calculations. A superimposed subset of the best refined structures yielded a medium resolution (backbone atoms' root mean square distance of 2.5 Å) tertiary conformation. The structure consists of three strands of anti-parallel beta sheet bound by three disulfide bonds and connected by short loops and turns, including a modified type VI (*cis*-proline) turn. The N-terminal region is not well defined due to a paucity of constraints. Myotoxin *a* exists as an equilibrium mixture of two forms in a 4:1 ratio, as evidenced by reverse-phase high performance liquid chromatography. Additionally, each form contains a small amount of what appears to be myotoxin *C. v. viridis*-2. Equilibrium of both forms is established within one hour from a single, isolated form at 25°C, but isolation at 2°C reduces the rate of interconversion. The existence of both chemical and conformational heterogeneity has produced complex NMR spectra with many peaks that can not be unambiguously assigned. These ambiguities have limited the number of distance constraints obtained and precluded the determination of a more highly defined tertiary structure. Purification of myotoxin *a* by affinity chromatography and low temperature separation of conformers should greatly facilitate generation of well-refined, highly-converged, accurate structures.

Dedicated to my departed father, Robert,

who was a source of inspiration,

and to my mother, wife and children

for their continuing encouragement.

ACKNOWLEDGEMENTS

I thank the Department of Chemistry and Biochemistry for assisting me in making the most of my opportunities. In particular, I thank Janet Snell for helping to smooth the administrative work and Dr. Allan Bieber for being a supportive mentor. I also thank Dr. Ronald Nieman for many hours of personal instruction and guidance.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER I

INTRODUCTION

The prairie rattlesnake (*Crotalus viridis viridis*), a subspecies of the western rattlesnake, inhabits a region of the central United States from western Montana eastward through the Dakotas, south through western Texas and into northeastern Arizona. This habitat extends slightly into Canada and Mexico but does not include much of the agricultural regions of Nebraska and Kansas. The presence of *C. v. viridis* in Arizona has been attributed to Indians bringing these snakes to Hopi villages for snake dances (Klauber,1982).

C. v. viridis may grow to a length of about 1.75 meters and can strike to a distance half its length at over 3 meters per second (Klauber,1982), envenoming the target. The snake, when milked, will yield approximately 44 mg (dry weight) or more of venom. With its venom's intravenous toxicity (LD_{50} in mice, mg/kg) of 1.0 -1.6 and intraperitoneal toxicity of 2.0-2.3, the adult prairie rattlesnake holds over a thousand lethal doses (20g mice) in its venom glands (Glenn & Straight, 1982).

An envenomed adult human may experience localized stinging, numbness, tingling at the extremities, nausea, localized swelling and discoloration, extreme pain, faintness and coma. A child victim may additionally exhibit hypertonic muscles and convulsions (Klauber,1982).

Myotoxins. While the venom of *C. v. viridis* contains many components that elicit various systemic and localized responses in victims, the focus of this study is on myotoxin *a*, a small, 42-residue protein that induces myonecrosis upon envenomation. The biological purpose has been suggested as the limitation of flight of prey and promotion of death caused by the paralysis of the limbs and diaphragm, respectively (Ownby et al., 1988; Griffin & Aird, 1990). In humans, the myonecrosis can cause permanent damage, leading to the loss of extremities (Tu, 1991). Myotoxin *a* is one of two myotoxic components in the

venom of *C. v. viridis*. The other, viriditoxin, is a high molecular weight protein that exhibits myotoxic effects secondary to its hemorrhagic effects (Gleason et al., 1983). In contrast, myotoxin *a* is of small molecular mass (M_r 4824) (Griffin & Aird, 1990) and contains only a non-enzymatic myotoxic activity (Mebs & Ownby, 1990).

Light microscope studies of mouse skeletal muscle tissue after i.m. injection of myotoxin *a* revealed vacuolation attributed to enlargement of the sarcoplasmic reticulum vesicles within 3 hours. By 12 hours, examination showed not only a continued dilation of the sarcoplasmic reticulum but also a swelling of the perinuclear space. By 48 hours, the highly enlarged sarcoplasmic reticulum had degraded into several smaller vesicles, mitochondria had enlarged, and the myofibrils had begun to disintegrate. After 72 hours, the myofilaments were completely disorganized and the cells were necrotic (Ownby, 1982; Mebs & Ownby, 1990).

A mode of action for myotoxin *a* has been suggested as an inhibition of the Na^+/K^+ ATPase, causing an influx of Na^+ with its solvating water swelling the sarcoplasmic reticulum and then the entire cell until fatally disrupted (Ownby, 1982). Electrophysiological investigations of myotoxin *a* on mouse and rat skeletal muscles revealed a reduced (-80mV to -60mV) resting membrane potential that was reversed by tetrodotoxin (Na^+ channel inhibitor) or low [Na^+], enhanced by ouabain (Na^+-K^+ ATPase inhibitor) or low [Cl^-], and unaffected by [K^+]. These findings suggest that the direct target is the sarcolemma's Na^+ channel, with myotoxin *a* serving to increase Na^+ permeability (Hong & Chang, 1985). Incubation of frozen, sectioned human muscle tissue with horseradish peroxidase-conjugated myotoxin *a* showed binding to the sarcoplasmic reticulum rather than to the sarcolemma (Tu & Morita, 1983).

Myotoxin *a* and its N- and C-terminal fragments have been shown to bind to Ca^{++} ATPase and prevent Ca^{++} uptake in isolated sarcoplasmic reticulum vesicles (Baker et al., 1992; Utaisincharoen et al., 1991), suggesting that myotoxin *a* may act as an inhibitor of the Ca^{++} ATPase of the sarcoplasmic reticulum membrane. However, Engle et al. (1983) found no change in sarcoplasmic reticulum vesicles' Ca^{++} uptake or release when treated with myotoxin II from *C. v. concolor*, a highly homologous myotoxin with similar histological effects (Ownby et al., 1988).

Previous Structural Studies. Myotoxin *a* belongs to a unique, yet highly homologous family of proteins whose members are all small myotoxins from snake venom (Bieber et al., 1987). These members include myotoxins from *C. durissus terrificus*, *C. adamanteus*, *C. scutulatus scutulatus*, *C. v. concolor*, *C. v. helleri*, and several forms from *C. v. viridis*, including myotoxin *a*, that exhibit sequence microheterogeneity from one another (see Figure 1). Many residues are completely conserved and substitutions are limited to a few, often conservative, substitutions.

The cysteine residues are almost completely conserved. Alternative disulfide bond arrangements have been reported for crotamine, including C4-C37 / C11-C36 / C18-C30 (Conti & Laure, 1988) and interchain disulfide links to form homodimers and, perhaps, polymers up to a hexamer (Teno et al., 1990). Nevertheless, light microscope studies of myotoxin *a* and crotamine on mouse skeletal muscle cells have shown cellular damage of the same histology (Cameron & Tu, 1978). Recent unpublished results from Bieber et al. confirm the disulfide arrangement of myotoxin *a* as reported by Fox et al. (1979) and show no evidence for covalent dimerization of myotoxin *a*.

Toxin	Source	Amino Acid Sequence	Reference
		1 5 10 15 20 25 30 35 40 45	
myotoxin *a*	*C. v. viridis*	YKQCHKKGGHCFPKEKICIPPSSDLGKMDCRWKWKCCKKGSG	Fox et al., 1979
myotoxin	*C. s. scutulatus*	YKQCHKKGGHCFPKEKICIPPSSDLGKMDCRWKWKCCKKGSG	Bieber et al., unpub.
crotamine	*C. d. terrificus*	YKQCHKKGGHCFPKEKIC**L**PPSSD**F**GKM**S**CRW**R**WKCCKKGSG	Laure, 1975
myotoxin I	*C. v. concolor*	YK**R**CHKK**E**GHCFPK**TV**IC**L**PPSSD**F**GKMDCRWKWKCCKKGS**VN**	Bieber et al., 1987
myotoxin II	*C. v. concolor*	YK**R**CHKKGGHCFPKEKIC**T**PPSSD**F**GKMDCRWKWKCCKKGS**VN**	Bieber et al., 1987
peptide C	*C. v. helleri*	YK**R**CHKKGGHCFPK**TV**IC**L**PPSSD**F**GKMDCRWKWKCCKKGS**VN**	Meada et al., 1978
CAM-toxin	*C. adamanteus*	YK**R**CHKKGGHCFPK**TV**IC**L**PPSSD**F**GKMDCR**W**R**W**KCCKKGS**VNNA**	Samejima et al., 1991
viridis-2	*C. v. viridis*	YK**R**CHKK**E**GHCFPK**TV**IC**L**PPSSD**F**GKMDCRWKWKCCKKGS**VNNA**	Griffin & Aird, 1990
viridis-3	*C. v. viridis*	YK**R**CHKKGGHCFPK**TV**IC**L**PPSSD**F**GKMDCRWKWKCCKKGS**VNNA**	Griffin & Aird, 1990
viridis-4	*C. v. viridis*	YK**R**CHKK**E**GHCFPK**TV**IC**L**PPSSD**F**GKMDCRWKWKCCKKGS**VN**	Griffin & Aird, 1990
viridis 2	*C. v. viridis*	YKQCHKKGGHCFPKEKICIPPSSD**K**GKMDCRWKWKCCKKGSG	Aird et al., 1991
viridis 3	*C. v. viridis*	YKQCHKKGGHCFPKEKICIPPSSDLGKMDCRWKWKCCKKGSG	Aird et al., 1991
viridis 4	*C. v. viridis*	YKQCHKKGGHCFPKEKICIPPSSD**K**GKMDCRWKWKCCKKGSG	Aird et al., 1991
viridis 5	*C. v. viridis*	YKQCHKKGGHCFPKEKIC**T**PPSSD**F**GKMDCRW**R**WKCCKKG	Aird et al., 1991
viridis 6	*C. v. viridis*	YKQCHKKGGHCFPKEKICIPPSSDLGKMDCRWKWKCCKKGSG	Aird et al., 1991
pCM45B	*C. d. terrificus*	YK**R**CH**I**KGGHCFPKEKICIPPSSD**F**GKMDC**PWRR**KCCKKGSG	Smith & Schmidt, 1990
pCM31B	*C. d. terrificus*	YK**R**CH**I**KGGHCFPKE**L**ICIPPSSD**F**GKMDC**PWRRKSL**KKGSG	Smith & Schmidt, 1990
pCM26B	*C. d. terrificus*	YK**R**CH**I**KGGHCFPK**G**KICIPPSSD**F**GKMDC**PWRR**KCCKKGSG	Smith & Schmidt, 1990
pCM20B	*C. d. terrificus*	YKQCHKKGGHCFPKEKICIPPSSD**F**GKMDCR**W**R**W**KCCDDGSG	Smith & Schmidt, 1990

FIGURE 1: Amino acid sequences of myotoxic peptides. Microheterogeneity exists in the sequences of myotoxin II (E15T; K16V; T19L; K33R), *viridis 2* (I19T; K25F), *viridis 3* (L25K), and *viridis 4* (K25L). The sequences of Smith & Schmidt are those encoded by cDNA and represent the microheterogeneity found in directly sequenced crotamine (K6I; R31P; W34R). Additionally, Aird et al. (1991) cite a personal communication reference to a highly homologous, 42-residue myotoxin, toxin E, from *C. h. horridus*. Bold letters indicate residues not in concensus with myotoxin *a*.

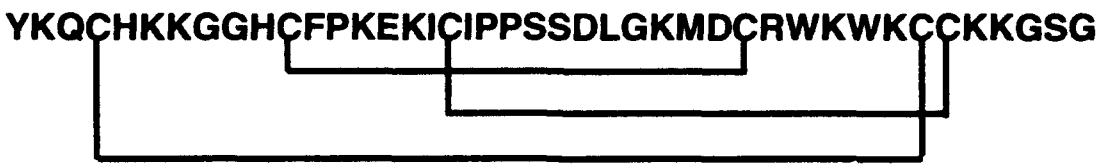


FIGURE 2: Disulfide bond arrangement of myotoxin *a* (Fox et al., 1979).

Initial circular dichroism (CD) studies suggest that myotoxin *a* degrades to a random coil structure upon reduction and alkylation of the disulfide bonds. Light microscope studies revealed a loss of biological activity with this chemically modified form, leading to the conclusion that the disulfide bonds were necessary for myotoxin *a*'s toxicity (Cameron & Tu, 1977). The disulfide bridges alone appear to impart great conformational constraints on the native structure. Note, however, that Baker et al. (1992) showed an equal inhibition of Ca^{++} uptake in isolated sarcoplasmic reticulum vesicles from N- and C-terminal peptide fragments containing no disulfide bonds as from native myotoxin *a*. The link between interactions of myotoxin *a* on Ca^{++} ATPase and observed myonecrosis has not been firmly established.

Predictive and experimental methods led to the conclusion that the secondary structure of myotoxin *a* appeared to be mostly β-sheet with little or no α-helix. Bailey et al. (1979) employed a modified Chou-Fasman secondary structure prediction technique to yield figures of 14% α-helix / 57% β-sheet and 64% α-helix / 47% β-sheet (overlapping) when using original and revised parameters, respectively. CD spectroscopy indicated no α-helix but instead indicated β-sheet and β-turns. Laser Raman analysis gave results of no α-helix, 73% β-sheet, and 27% random coil.

Using a predictive method based on hydrophobicity, Henderson and Bieber (1987) suggested the presence of a 14-residue N-terminal α-helix on a

structure otherwise composed of β-sheet. The pH-dependent ^{1}H-NMR shifts of Y1, H5, and H10 suggested that these residues are in close proximity of each other, as in a N-terminal helix. Furthermore, the pH titration shifts suggested that the protonation of one or both histidine residues (H5 & H10) causes the N-terminal region to become exposed to solvent, which would be consistent with a helix (at physiological pH) to random coil (at low pH) transition. The side chain of Y1 showed a coupling pattern that suggested free ring rotation (Henderson et al., 1987). It is interesting to note that this Y1 is completely conserved among all members of this family of myotoxins and is essential for the activity of myotoxin *a* (Hayes, 1984).

Henderson (1986) completed the assignment of the aromatic residue side chains as well as that of three singly occurring residues (L25, M28 & R31) using one- and two-dimensional techniques. Using 400 MHz ^{1}H-NMR two-dimensional spectra (COSY, NOESY and RELAY), Murchison (1989) was able to make sequence-specific assignments for the NH, $C^{\alpha}H$, and $C^{\beta}H$ peaks of ca. 50% of the residues in myotoxin *a*. The sequential assignment technique (Billeter et al., 1982; Wüthrich, 1986) led to the successful assignment by Murchison of only the region R31-K38 . The amino acid types could only be discerned as being G (AX), AMX or long-chain spin systems. The use of a main chain directed search algorithm (Englander & Wand, 1987) found NOEs indicative of anti-parallel β-sheet between the regions 4-5 / 30-38 / 17-18. This MCD method also suggested a type II turn at residues K7-G8-G9 but found no evidence for α-helix anywhere in the structure.

Complexity of Spectra. Both Henderson (1986) and Murchison (1989) found the anomaly of too many peaks in the NMR spectra for the number of protons in a single form of myotoxin *a*. These excess peaks led to ambiguities

that prevented Murchison from assigning C11-I17 and other regions. Despite the complexity of spectra, subsequent work by Murchison and Nieman (unpub.) using 500MHz ^{1}H-NMR 2D experiments led to the assignment of peaks to nearly all protons in myotoxin *a*. Ambiguities and a lack of assignments still existed in the N-terminal region, which may not be structurally well defined under the conditions studied. Other incomplete assignments existed for some side chains, and many peaks in the spectra remained unassigned. No definitive attempt had been made to model an initial set of constraints from NOESY peak volumes into three-dimensional structures.

The heterogeneity of the myotoxin *a* preparation resulting in the complexity of NMR spectra has also manifested itself as twin peaks on RP-HPLC separations (Murchison, 1989). The explanation for the excess NMR peaks has been suggested as a similar protein contaminant that co-purified with myotoxin *a* (Henderson, 1986; Murchison, 1989), aggregation (Henderson, 1986), and/or isomerization, to include various possible disulfide bond arrangements or proline *cis-trans* isomerization at one of the three proline residues (Murchison, 1989).

Further investigation of the heterogeneity by Misra (1991) revealed that the myotoxin *a* preparation gave two peaks on RP-HPLC in a ca. 4:1 ratio of areas, with the minor peak eluting first. Injection of either peak, after drying and redissolving at ambient temperature, resulted in the appearance of both peaks in about the same ratio of areas. Carboxypeptidase Y treatment of each original peak's fraction to remove the C-terminal residue and subsequent amino acid analysis of those residues revealed the presence of glycine and alanine in a ca. 4:1 ratio in each peak. The glycine would correspond to G42 of myotoxin *a*, and the alanine would likely correspond to A45 of *viridis*-3 (Griffin & Aird, 1990). The implication is that both isomerization and chemical microheterogeneity are

present in the myotoxin *a* preparations and are responsible for the additional peaks present in NMR spectra.

Myotoxin *a* provides a suitable subject for study by NMR as it is small, highly soluble, available in sufficient amounts, and has not been reported as crystallized for x-ray studies. Though pharmacological research on myotoxin *a* has not been prevalent (Stocker, 1990), much work has been done toward neutralizing myotoxic activity (Ménez, 1991) and determining its site of action (Tu, 1991). Myotoxin *a* presents an intriguing problem in solving its structure-function relationships since neither have yet been well defined. The complexity of spectra makes this process a challenge.

CHAPTER II

MATERIALS AND METHODS

Purification. Myotoxin *a* was purified from *Crotalus viridis viridis* venom (Laser Lab, Salt Lake City) using essentially the method of Henderson (1986). One gram of dry crude venom was dissolved in 5 ml of 0.1M KCl/0.05M Tris buffer (pH 9.0) and centrifuged (Sorvall RC-5B Refrigerated Superspeed Centrifuge) 5 min at 1000 rpm at 4°C. The pellet was resuspended in 5 ml of the same buffer and the suspension was again centrifuged under the same conditions. The collected supernatant solutions from both centrifugations were pooled. This sample was run (Buchler peristaltic pump) onto a Fractogel TSK CM650S (EM Sciences) carboxymethyl cellulose cation exchange column (2.5 x 37.5 cm; Pharmacia) at 2°C in 0.1M KCl/0.05M Tris buffer (pH 9.0) at a flow rate of 1.5 ml/min with a salt gradient from 0.1 to 1M KCl. Absorbance at 280 nm was measured on an Instrumentation Specialties Co. (ISCO) Type 6 Optical Unit, amplified by an ISCO Model 1133 Multiplexer-Expander, and chart recorded on an ISCO Model UA-5 Absorbance/Fluorescence Monitor. Fractions were collected automatically (4 min/tube) on a Gibson FC-100 Micro Fractionator. Most proteins came off in the first major peak, but myotoxin *a*, with a higher pI, eluted much later, in the second major peak. The fraction containing myotoxin *a* was ultrafiltered in a 43 mm Amicon Model 52 concentrator using a Diaflo YM2 membrane filter (Amicon) at 2°C with 55psi N_2 through several dilution/concentration cycles. The sample was lyophilized (Virtis Freeze Dryer) and kept desiccated at -20°C until used.

NMR Sample Preparation. Lyophilized myotoxin *a* was dissolved to ca. 3-4 mM in 85% H_2O/15% D_2O with a small amount of 3-(trimethyl)propionic-2,2,3,3,-d_4 acid (TSP) to a total volume of ca. 0.8 ml. A typical sample was prepared by dissolving 15.3 mg myotoxin *a* in 0.86 ml 85% H_2O/15% D_2O (Aldrich). To this solution was added 50 µl of 0.075% (w/w) TSP (Aldrich) in 90%

H_2O/10% D_2O. The pH was monitored on a Radiometer Copenhagen PHM 84 Research pH meter with a Lazar PHR-146 Micro Combination electrode and adjusted to 3.5 with 1M and 0.1M HCl. A 50 µl aliquot was diluted into 950 µl H_2O, and its absorbance at 280 nm was measured on a Varian DMS 100S UV Visible Spectrophotometer. An extinction coefficient, ε_{280}=2.27 $mg^{-1}ml \cdot cm^{-1}$ (Allen et al., 1986) was employed to calculate the sample concentration.

NMR Acquisition Parameters. ^{1}H-NMR spectra were acquired on a Varian Unity 500 spectrometer controlled by a Sun 4/260 workstation running Varian's VNMR software. The spectrometer was operated at 499.843 MHz with a sweep width of 6199.6 Hz. Unless otherwise stated, data was acquired at 25°C. One-dimensional spectra (64 transients) were acquired with 32 K (32,768) points and referenced to TSP. Low-power continuous pre-irradiation was employed to suppress the water resonance. Double quantum filtered correlated spectroscopy (DQFCOSY) (Rance et al., 1983) experiments were normally acquired with 4 K points, 48 scans, and 800 increments. Total correlation spectroscopy (TOCSY) (Bax & Davis, 1985) spectra were normally acquired with 2 K points, 32 scans, and 600 increments with various mixing times from 20 to 100 ms. Nuclear Overhauser enhancement spectroscopy (NOESY) (Kumar et al., 1980) experiments were usually acquired with 2 K points, 96 scans, and 512 increments with various mixing times from 50 to 350 ms. Spectra were acquired with phase sensitive detection using the hypercomplex method (States et al., 1982).

In a software variation of the above experiments, both the decoupler and transmitter signals are synthesized from the transmitter board, rather than separately from the decoupler and transmitter boards. This modification proved to effectively suppress the water signal and allowed a greater sensitivity closer to

the water resonance, due to greater phase coherence between presaturation and observe transmissions. These experiments have sometimes been referred to as "transmitter-NOESY," "transmitter-TOCSY," etc. (TNNOESY, TNTOCSY, etc.) experiments. However, the pulse sequences are identical to those of the original experiments.

Spectral Assignment Strategies. Spectra were initially Fourier transformed on the Sun 4/260 workstation under VNMR. Parameters and data were passed to a Silicon Graphics Indigo R3000 workstation and subsequently converted into the format of Felix 2.05 (Hare Research, Inc.). The spectra were Fourier transformed in Felix using a sine square windowing function shifted $\pi/3$ to $\pi/6$ and were zero filled to a typical digital resolution of 3.02 Hz/point in F2 and 3.02 Hz/point in F1. The resultant 2 K x 2 K real matrices were inspected, and a suitable cutoff threshold was selected that minimized signal loss while reducing noise to a workable level. The size of the matrices were reduced by saving only data points above the threshold values. The "squeezed" matrices resulted in a large file space savings while maintaining nearly all usable information. These matrices were used to make spectral assignments and distance constraints in Felix, running on the Indigo or a Silicon Graphics Iris 4D/80GT workstation.

The general strategy for making spectral assignments followed Wüthrich (1986). First, the chemical shifts and dimensions of all peaks in the fingerprint (NH-$C^{\alpha}H$) region (ca. 3.5-6.5 ppm in one dimension and 6.5-11 ppm in the other dimension) were entered into a Felix database. Peak centers and widths were determined interactively using the Felix graphical interface. Each crosspeak resulting from the scalar coupling of the NH and $C^{\alpha}H$ protons within the same residue was given a unique numerical assignment. For example, the database entry for one such COSY fingerprint peak might show nh23 as the assignment in

one dimension and ca23 as the assignment in the other dimension. The corresponding crosspeaks on the opposite side of the diagonal in the COSY spectra received the same arbitrary numerical assignment.

Second, peaks were picked in the $C^{\alpha}H$-$C^{\beta}H$ region (ca. 2.0-3.5 ppm in one dimension and 3.5-6.0 ppm in the other dimension). The assignments of peaks in this region were then brothered across the diagonal to corresponding NH-$C^{\alpha}H$ peaks in the fingerprint region (see Figure 3). Peaks in the spectra that has similar chemical shifts along the $C^{\alpha}H$ resonance axis were displayed in expanded, aligned tiles. Peaks which were unambiguously aligned with the NH-$C^{\alpha}H$ peak were assigned as $C^{\alpha}H$-$C^{\beta}H$ peaks bearing the same arbitrary numerical assignment. In cases where more than one NH-$C^{\alpha}H$ peak or more than two $C^{\alpha}H$-$C^{\beta}H$ peaks fell along the same line, an unambiguous brothering could not be made. The $C^{\alpha}H$-$C^{\beta}H$ peaks in such a case would remain unassigned for the time being.

Third, all peaks in the TOCSY spectrum of a medium mixing time (60-80 ms) were picked into a separate database. The NH-$C^{\alpha}H$ peaks in the TOCSY spectrum were identified by the peaks in the corresponding locations on the COSY spectrum and given the same arbitrary assignments. Side chain TOCSY peaks with the same NH chemical shift were brothered region from 6.5-11 ppm in one dimension and 0.5-6.0 ppm in the other dimension (referred to hereafter as the NH-$C^{\delta}H$ region). Whenever possible, $C^{\alpha}H$-$C^{\beta}H$ peak assignments from the COSY spectrum would be transferred to NH-$C^{\beta}H$ peaks in the TOCSY spectrum with the same $C^{\beta}H$ shifts and which were brothered to the same $C^{\alpha}H$ shift. The remaining TOCSY peaks along the NH resonance were initially assigned as C? and the arbitrary number of the residue.

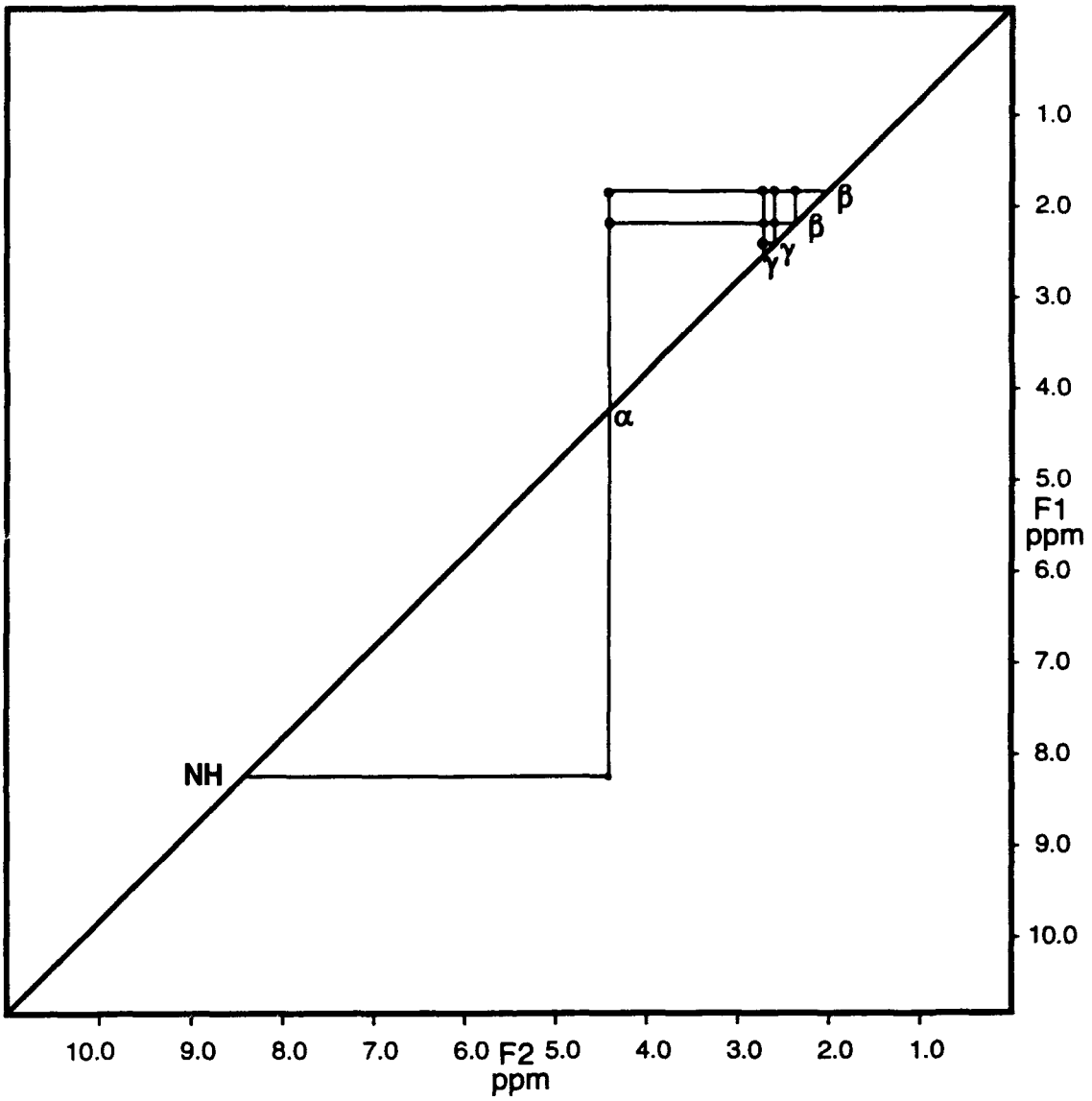


FIGURE 3: Schematic diagram of the connectivity between NH-C$^{\alpha}$H, C$^{\alpha}$H-C$^{\beta}$H, and C$^{\beta}$H-C$^{\gamma}$H peaks in a COSY spectrum for a glutamate residue. Another symmetry-related pattern exists across the diagonal (not shown).

Fourth, the pattern of peaks from each given NH shift in the NH-C$^{\delta}$H region was followed to where it was repeated at the C$^{\alpha}$H, C$^{\beta}$H, and other side chain proton shifts (see Figure 4). The arbitrary assignments were correspondingly brothered. This process was done in an iterative manner along with picking and identifying the corresponding side chain peaks (past C$^{\beta}$Hs) in the COSY spectrum and transferring the TOCSY assignments back to these peaks. The iterative nature of the process allowed resolution of some ambiguities and identification of some of the outer side chain protons as C$^{\gamma}$H, C$^{\delta}$H, etc. Additionally, use was made of the TOCSY spectra at different (usually 100 ms) mixing times to better reveal particular peaks in certain regions of the spectrum. Brothered side chains were identified as either G (AX), AMX, L, R, I, S (AMX with C$^{\beta}$H peaks downfield of 3.5 ppm), P (absent in the NH-C$^{\delta}$H region; identified last after assigning most of the C$^{\alpha}$H-C$^{\delta}$H region, 0.5 to 6.0 ppm in both dimensions), or long chain (Wüthrich, 1986).

Fifth, peaks in the aromatic region (ca. 6.0-8.0 ppm in both dimensions) of the COSY were picked. Assignments of types of residues were made based upon the characteristic spin coupling patterns. In the case of aromatic side chain peaks that could be unambiguously assigned, these assignments were then transferred to the corresponding peaks in the TOCSY. Where possible, the aromatic side chains were brothered to their corresponding C$^{\alpha}$H peaks in the TOCSY spectrum.

Sixth, all peaks in the NOESY spectrum of a medium mixing time (100 ms) were picked without any initial assignments. Assignments were transferred from the corresponding TOCSY peaks. The TOCSY crosspeaks represent all the intraresidue interactions. The NOESY crosspeaks represent all through-space interactions of protons within ca. 5 Å of each other. Therefore, after

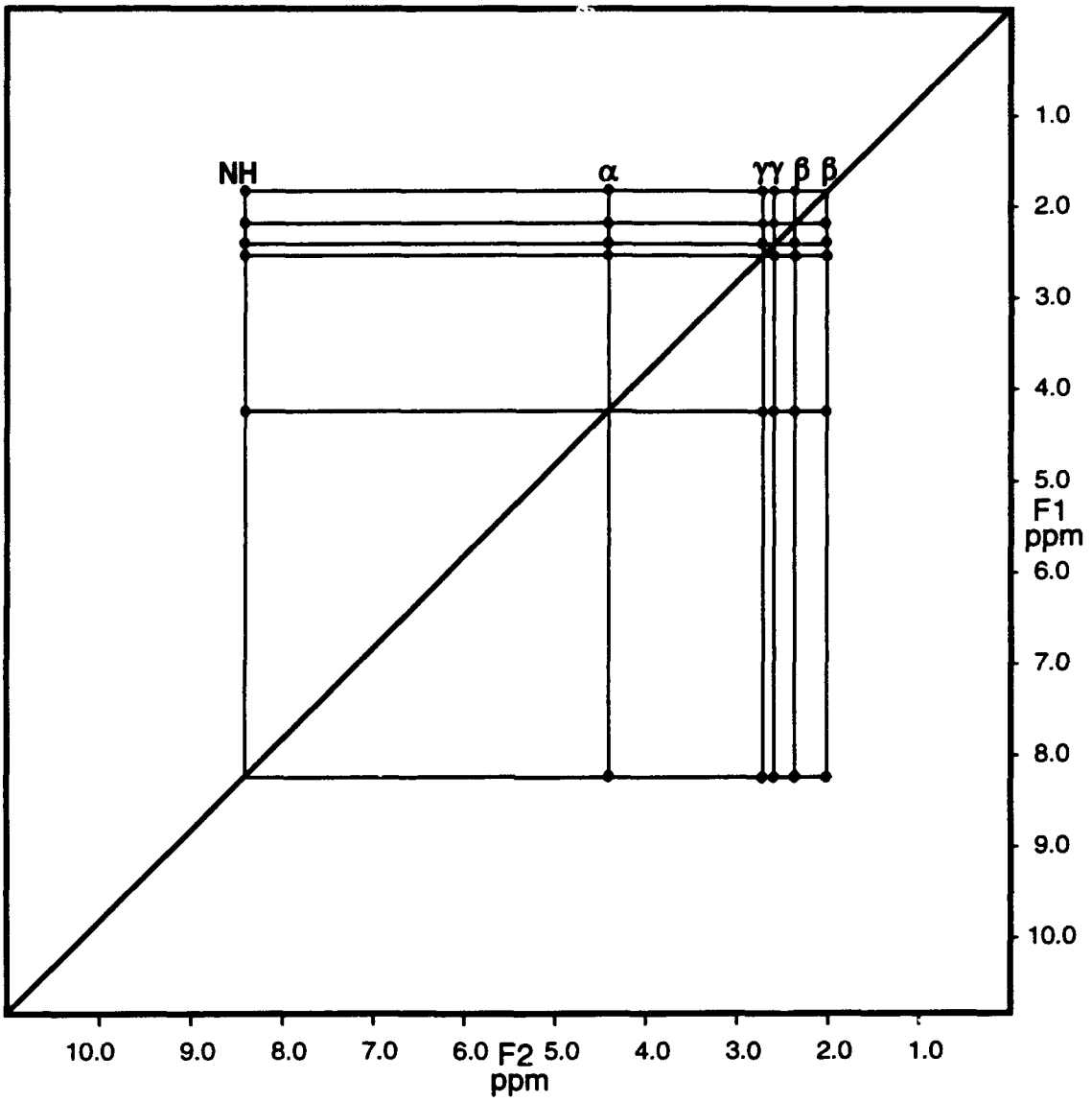


FIGURE 4: Schematic diagram of the recurring connectivity in a TOCSY spectrum for a glutamate residue. For long side chains, not all peaks are visible at each resonance, depending on mixing time.

transferring the assignments from the TOCSY to the NOESY spectrum, the remaining unassigned NOESY crosspeaks represented interresidue interactions. Wüthrich (1986) has shown that for allowable ϕ,ψ angles, NH_i-$C^{\alpha}H_i$ (intraresidue) crosspeaks and $C^{\alpha}H_i$-NH_{i+1} (sequential, interresidue) crosspeaks will occur in the NOESY spectrum (given the proper acquisition parameters). The sequential connectivity of the residues in the spectrum is given by connecting a NH_i-$C^{\alpha}H_i$ peak to the NH_{i+1}-$C^{\alpha}H_i$ peak to the NH_{i+1}-$C^{\alpha}H_{i+1}$ peak to the NH_{i+2}-$C^{\alpha}H_{i+1}$ peak and so on. However, the chain of such sequential connectivities breaks where a proline (no NH) is present in the sequence or where more than two peaks are in the fingerprint region are aligned. Since the peaks in this region are not confined to NH_i-$C^{\alpha}H_i$ and $C^{\alpha}H_i$-NH_{i+1} peaks, many NH_i-$C^{\alpha}H_j$ peaks (interresidue, nonsequential) may exist, precluding unambiguous assignment of sequential connectivities. Wüthrich (1986) suggested that basing the sequential connectivities solely upon these connections would likely yield only about half correct results. Ambiguities were resolved and fingerprint region connectivities were verified by using the NH-NH region (6.5-11 ppm in both dimensions) of NOESY spectra obtained at different pH and temperatures (usually samples of different pH; here, NH shifts are greater than $C^{\alpha}H$ shifts).

By combining the sequential connectivities with the types of spin systems of the side chains, connected peaks were then married up with their uniquely corresponding segments of the primary sequence in order to make the sequence-specific assignments. Conversion of the Felix database arbitrary numerical assignments to the sequence specific assignments was done by writing out the entities for each spectrum, filtering each written file through a UNIX sed-based script conversion file (see Appendix B), and reading these new files back into their appropriate entities.

NOE Distance Constraints. A simplistic approach to deriving distance constraints from NOE data is to integrate the volumes of the peaks at a given medium mixing time (ca. 100 ms) that allows for peak buildup but minimizes effects of spin diffusion. Volume integrals were calibrated to a peak whose represented distance is known in the structure. Wüthrich (1986) suggested analyzing for secondary structural features first, then calibrating from the known distances in such structures. The derived distance would serve as an upper limit distance constraint. The lower distance constraint would be set to the sum of the van der Waals radii of the two protons. A more thorough approach calculates distance constraints from the buildup of the peaks from several spectra acquired at different mixing times. However, distance constraints must still be calibrated to a known structural distance. Felix provides an automated tool for doing this type of determination. The NOESY peak boxes that have been picked over the peaks of a NOESY at a single mixing time are sequentially laid over the matrices of the NOESY spectra at different mixing times. The peak volumes are integrated for each spectrum. To generate a table of calculated upper and lower distance constraints, a peak's volume integral and its corresponding distance for calibration were supplied.

One initial set of constraints was determined by Nieman (unpub.) by manual segregation of peaks into weak, medium weak, medium and strong based on their integrated volumes. The peaks were then assigned upper limit distance constraints of 5, 4, 3.5 and 2.8 Å, respectively. The lower distance constraints were set at 3 Å for weak peaks and the sum of the van der Waals radii for all other peaks.

A separate set of 311 distance constraints was determined from the first group of myotoxin *a* spectra to be completely analyzed in Felix. A 200 ms

TNNOESY spectrum of 3.4 mM myotoxin *a* at pH 3.5 in 85% H_2O/15% D_2O at 25°C provided a single set of peaks whose volumes were integrated. These volumes were calibrated against aromatic protons of fixed distance and converted into upper limit distance constraints by Felix.

Multiple NOESY spectra at various mixing times were not used because they were taken without transmitter presaturation and at different pH. Therefore, these spectra at multiple mixing times had lower sensitivity and would not accurately overlay the TNNOESY spectrum due to shifted peaks. To use these spectra for determining volume buildup rates and subsequent distance constraints would have involved attempting to accurately follow all the shift changes. The chosen TNNOESY aligned with a 60 ms TNTOCSY and a TNDQCOSY of the same sample. These spectra were chosen to be fully analyzed under Felix, with qualitative rather than quantitative augmentation from other spectra of other similar samples, because they were the highest quality spectra of myotoxin *a* to date and provided the best data for making assignments.

The set of 311 distance constraints was manually screened to eliminate redundant constraints. Any constraints on residues Y1, K2, Q3, H5, and K6 were also eliminated because of a lower confidence in their assignments (see results). The resultant list of 134 experimental upper limit distance constraints (77 intraresidue, 29 interresidue sequential, 28 interresidue nonsequential) were rounded up to the next whole angstrom and manually entered into the required format for distance geometry calculations.

Distance Geometry Calculations. Initial distance geometry calculations and modeling were done with Dspace 4.0 (Hare Research, Inc.) running on a Silicon Graphics Iris 4D/80GT workstation. Dspace is an implementation of a

metric matrix distance geometry algorithm (Crippen, 1977; Wemmer, 1990), a method that has been shown to successfully determine protein structures in solution (Havel & Wüthrich, 1985; Williamson et al., 1985). The program, as supplied by Hare Research, came with incomplete tools for calculating protein structures. The functions provided in Dspace allow the user to build functioning macros to make the program perform an appropriate strategy of refinement. Non-functioning sample macros were included with the program.

Macros were written and revised to perform an effective refinement strategy (see Figure 5). The primary refinement macro, zipref.mac, and the other macros it calls are included in Appendix A. This macro "zippers" the protein from the N-terminal to the C-terminal starting with the refinement of the individual residues, repeating with a two-residue window, and repeating over and over with an ever increasing window size until the entire structure is being refined at once. Dspace calculates, on a recurring basis, a penalty function which is essentially a weighted sum of the differences between the allowable and the actual interatomic distances and angles in the structure at the time of the calculation. The lower the penalty function, the better the structure conforms to all the constraints, covalent (bond lengths and angles), steric (vdW radii), chiral (L-amino acids) and experimental (NOE-derived distance constraints). The refinement macro institutes simulated annealing when the refinement process fails to bring a segment to within a given penalty value, based on the window size.

Zipref.mac has many desirable features. By annealing and minimizing the difference between allowable and actual distances with single residues at first, the macro emphasizes correct local geometry, as very few experimental constraints are imposed on any given individual residue. Rigorous annealing in

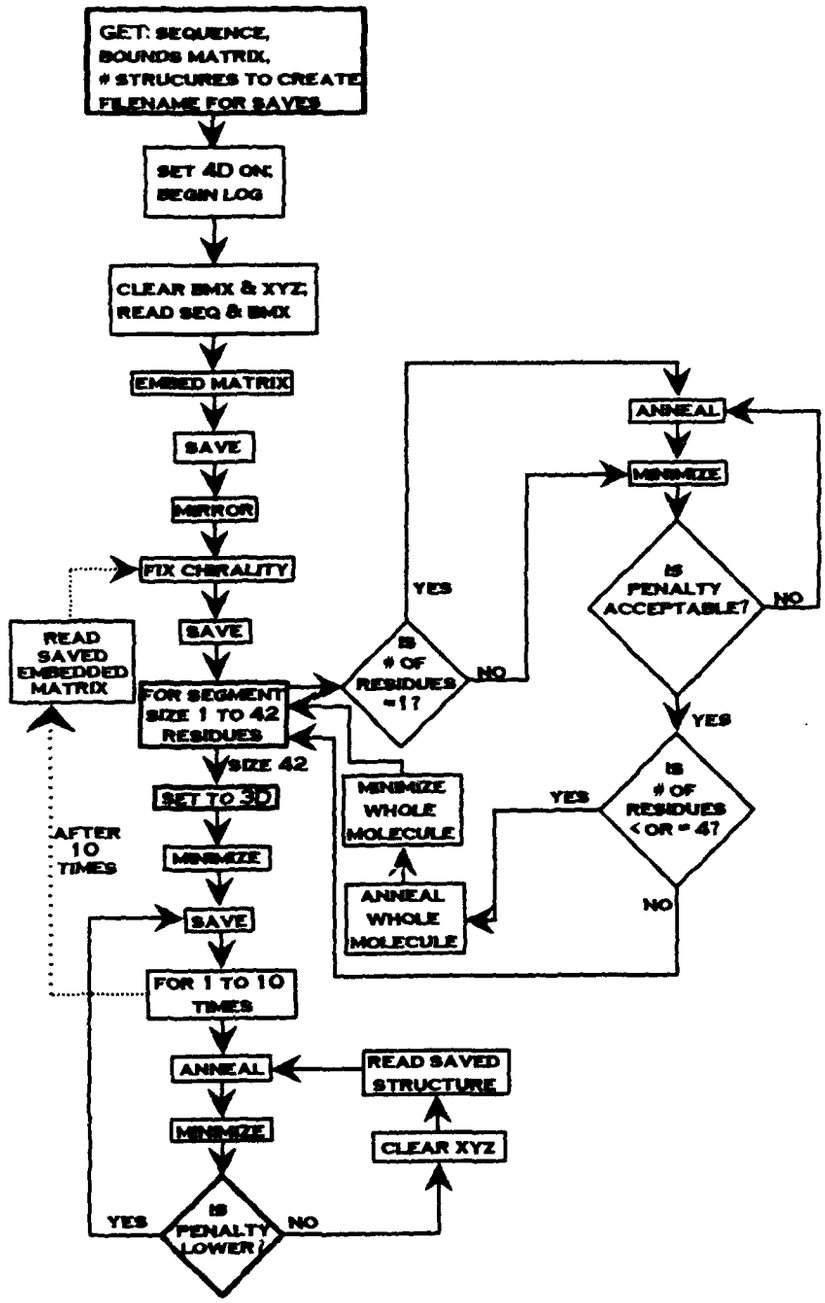


FIGURE 5: Logic flow diagram of the zipref.mac refinement macro for Dspace.

four-dimensional space takes place in the beginning of the refinement process, regardless of penalty, thus forcing Dspace to sample more conformational space. Limited conformational space sampling has been a noted limitation of this program (Metzler et al., 1989). The annealing and minimization of the whole molecule, periodically interspersed with the zippered refinement with an ever increasing window size, efficiently balances the effects of the local and global constraints. The macro constantly checks and corrects incorrect chirality at chiral centers while the program employs floating chirality at non-stereospecifically assigned prochiral centers (Weber et al., 1988). The macro refines both the original embed of atoms and its mirror image, generating both of a potential pair of "tertiary enantiomers" (structures with the same chirality within the primary structure but with essentially mirror image backbone folding). This means that an embed that leads to a well refined structure is not wasted on generating only the mirror image folding of the correct structure. Finally, zipref.mac, by following the progress of the penalty function, prevents acceptance of poor structures and tries to better the best structures.

Zipref.mac was employed with the original (Nieman, unpub.) constraints and versions thereof where suspect constraints were removed or modified.

Additional distance geometry calculations were performed with DIANA 1.14 (Distance Geometry Algorithm for NMR Applications) (Güntert et al., 1991a,b), which employs a variable target function distance geometry algorithm (Wemmer, 1990), a method also fully capable of determining protein structures in solution (Wagner et al., 1987). DIANA was compiled under FORTRAN-77 in UNICOS and run on a Cray X-MP supercomputer. DIANA was later compiled under FORTRAN-77 in IRIX and run on a Silicon Graphics Indigo R3000 computer. DIANA employs pseudoatoms to accommodate non-stereospecific

assignments at prochiral centers (Wüthrich et al., 1983). The default refinement strategy was used on both the original (Nieman, unpub.) and the TNNOESY-derived sets of distance constraints and modifications thereof.

Visualization and Evaluation. Structures generated in Dspace were visualized within Dspace on the Silicon Graphics Iris 4D/80GT workstation. Macros employed the program's abilities to superimpose structures, calculate root mean squared deviations (RMSDs), and display selected parts of the structure. Dspace has the ability to rotate line drawings in real time. RMSDs were calculated only to a single structure on which the others were then superimposed. To get a pairwise listing of RMSDs, a group of structures had to be repeatedly superimposed onto each member of that group.

Structures generated by DIANA were in the form of an atomic coordinate file that required format conversion to be imported into Quanta (Polygen) running on the Silicon Graphics Iris 4D/80GT workstation. The UNIX script conversion file appears in Appendix B. Quanta can superimpose structures, display selected atoms, and overlay a backbone tracing ribbon. Line drawings may be manipulated in real time, including in stereoview. DIANA outputs an overview file that contains a matrix of all pairwise RMSDs for structures with a target function below a chosen cutoff value. The overview file also contains a listing of repeatedly violated constraints as well as a listing of possible hydrogen bonds.

HPLC Separations. Reverse-phase high performance liquid chromatography (RP-HPLC) was performed on a BioRad HPLC Model 1330, controlled by BioRad software running on an Apple IIe computer. The sample of not more than 1 mg of myotoxin *a* in water was injected onto a Phenomenex Selectosil 5 C4 (250 x 10 mm; 5 micron) semi-preparative column. Samples were typically eluted with a gradient of 22-23% acetonitrile (Baker) in triple

distilled water with 0.01 M trifluoroacetic acid (Pierce) (ca. pH 2.0) over 25 minutes at a flow rate of 3.0 ml/min. Absorbance at 220 nm was measured with a LDC SpectroMonitor III flow cell and recorded and analyzed on a Spectra-Physics SP4100 Computing Integrator. Fractions were collected manually. Selected fractions were dried on a Speed Vac Concentrator. Low temperature separations were done at ca. 2°C by pre-chilling the solvents overnight in a cold room and keeping all solvents, samples, and collected fractions in ice baths during the course of the experiments.

GCG Analyses. The Genetics Computer Group (GCG) Sequence Analysis Software Package was run on a VAX 6000-430 system running VMS 5.5-2. Peptide sequence homology searching was performed using FastA with the SwissProt database of 20,024 sequences dated August 1992 using a word size of 2. A TFastA search of the GenEMBL database of 48,274 sequences dated September 1992 was performed with a word size of 2. A Motifs search was also conducted. PeptideStructure was used to perform automated Chou-Fasman (Chou & Fasman, 1974) and Garnier-Osguthorpe-Robson (Garnier et al., 1978) secondary structure predictions. The output of PeptideStructure was visualized with a PlotStructure one-dimensional plot. HelicalWheel was used to look at the alignment of side chains on a selected region of myotoxin *a* that showed the potential to be in an α-helix.

CHAPTER III

RESULTS AND DISCUSSION

General Quality of NMR Spectra. The TNCOSY, TNTOCSY and TNNOESY experiments produced the highest quality NMR spectra, especially close to the water resonance line. The COSY peaks were well shaped and easily picked, except in the case of multiple, overlapped peaks. The fingerprint region should have had at most 45 NH-$C^{\alpha}H$ peaks if all five glycine residues' pairs of $C^{\alpha}H$ peaks were distinguishable and if all 40 amide protons, including the N-terminal protons, were visible (42 residues + 5G - 3P + 1N-term. = 45). In reality, no NH-$C^{\alpha}H$ peak was found for Y1, so a more realistic expectation would be to see not more than 43 NH-$C^{\alpha}H$ peaks in the fingerprint region. At the lowest threshold of display, clearly 57 such peaks appeared in this region. There were many instances of weak peaks appearing very close to or partially overlapping strong peaks. The COSY spectrum became unusable due to severe overlap only in the upfield region occupied by $C^{\beta}H$-$C^{\gamma}H$ and further upfield crosspeaks (0.5-3.5 ppm in both dimensions). The high quality of the spectra allowed peak picking and brothering in both dimensions (on both sides of the diagonal). Since the resolution is not the same in both dimensions, the two sides of the spectrum were not exactly symmetrical. In fact, they were complementary to each other in that the splitting pattern was usually different in each dimension, allowing more accurate picking of the center of peaks and the deconvolution of overlapping peaks.

The TOCSY spectra were relatively similar despite using a range of mixing times from 20 to 100 ms. The best overall spectrum, at a mixing time of 60 ms, was used for brothering the side chain connectivities. The spectra of shorter mixing times lacked some peaks, while the 100 ms TOCSY had many peaks which were highly misshaped and overlapped. The 100 ms spectrum did contain a few more peaks than the 60 ms, and these were added to the spin

system assignments. The 80 ms TOCSY was very similar to the 60 ms TOCSY and probably could have been used just as effectively for making the spin system assignments. All of the TOCSY spectra suffered to some extent from misshapen peaks which were likely due to the slightly offset overlapping peaks resulting from the heterogeneity of the sample.

The 200 ms TNNOESY spectrum was superior to the other NOESY spectra and aligned properly with the TNTOCSY and TNCOSY spectra. In contrast to the TOCSY spectra, where spectra from different samples at different mixing times could be used in a complementary fashion, the various mixing time NOESY spectra from different samples could not be effectively used with the 200 ms TNNOESY. In COSY and TOCSY spectra, when peaks shifted slightly, the peaks could be easily correlated by following their spin systems. In the NOESY spectra, however, there were many unidentified peaks which led to many ambiguous connectivities. Therefore, to follow the shifting of peaks between NOESY spectra taken under slightly different conditions, the companion shifts in COSY and TOCSY spectra from the various conditions had to be first correlated. For these reasons, the 200 ms TNNOESY was used as the primary NOESY spectrum for making interresidue connectivities and deriving distance constraints. The NOESY spectra also suffered from asymmetric peaks, which were likely due to the heterogeneous nature of the sample.

The presence of chemical microheterogeneity and apparent isomerization exhibited itself in all spectra, both as excess peaks with distinctly separate chemical shifts and as peaks of lesser intensity in close proximity or partially overlapping stronger peaks. Excess peaks and shadow peaks could result from different causes, since extra peaks could result from peptides of slightly different composition or conformation. If chemical microheterogeneity and isomerization

are both present in ca. 4:1 ratios, the spectra will consist of peaks from at least three different peptide forms. The least populated form, the minor conformer of the minor sequence, would exist at low concentration and might not be visible, as would other minor forms which may be present. The major form, the primary conformer of myotoxin *a*, still accounted for the majority of peaks and could, in most cases, be unambiguously assigned. The heterogeneity of the sample had its most troublesome effect on interpretation of NOESY spectra where weak peaks could not be distinguished as strong NOEs from a minor form or weak NOEs from the major form.

Spectra obtained from different samples at slightly different pH and different temperatures proved to be most useful in sorting out ambiguities in connectivities and in finding peaks hidden by the suppressed water resonance in the 25°C spectra. Recording spectra at 25°C at slightly different pH caused the NH peaks to shift to some degree while most of the $C^{\alpha}H$ peaks did mot move significantly. In many cases, different sets of peaks aligned in different spectra. Since a connectivity must align in all spectra, the deconvolution of ambiguities was occasionally reduced to a process of elimination. The 50°C spectra proved invaluable in finding the NH-$C^{\alpha}H$ peaks of D29 and K35.

Assignment of Spin Systems. The first spin systems identified were the glycines, which each gave 2 NH-$C^{\alpha}H$ peaks in the COSY. The glycine peaks were more highly split than other peaks and displayed connectivity to $C^{\alpha}H$-$C^{\alpha}H$ peaks, which are unique to glycine residues. The remaining spin systems were identified by the complete side chain connectivities in the TOCSY (see Figure 6) and the complementary pattern in the COSY.

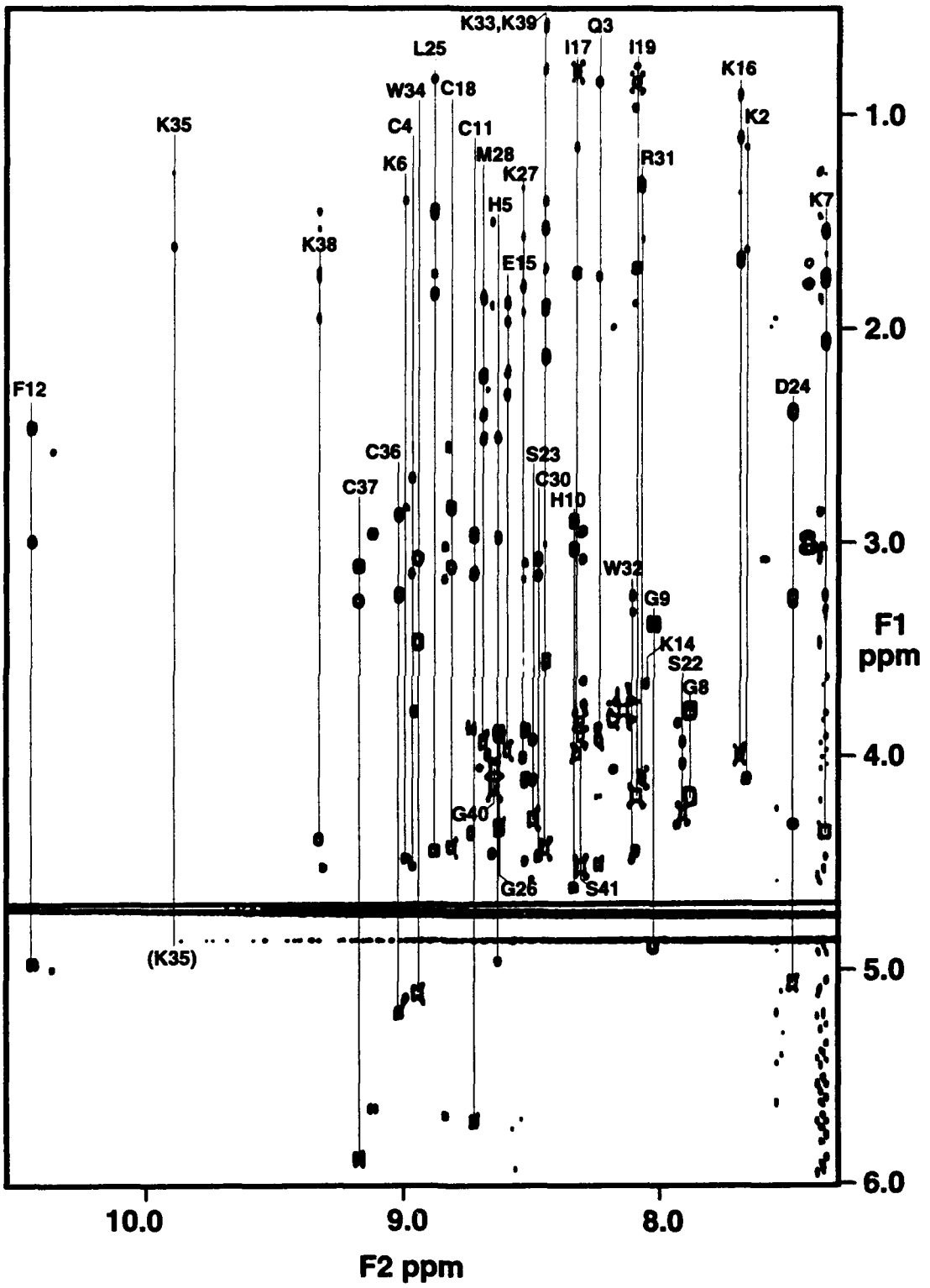


FIGURE 6: Side chain connectivities in TOCSY (60 ms) of myotoxin *a*, 3.4mM, pH 3.5 in 85% H_2O/15% D_2O at 25°C. D29 and K35 appear in 50°C spectra.

The two isoleucine residues were initially tentatively identified by their unique COSY pattern, though not all such peaks were resolved (see Figure 7). No other residue with a single $C^{\beta}H$ occurs in myotoxin *a*. To preclude misassignment due to coincidental $C^{\beta}H$ methylene shifts, both isoleucines' identities were also corroborated by their sequential connectivities.

Although leucine provides a unique COSY pattern, L25 was only initially identifiable as a long chain residue. The subsequent connectivity resolved the identity of this residue. The 10 lysine residues were identified as long chain residues and later confirmed as lysines by their connectivities. The overlap of peaks in the far upfield region (0.5-3.5) of the COSY made it impossible to completely follow the unique pattern of the lysine side chain all the way through their spin systems. The three other residues Wüthrich (1986) identifies as having unique spin systems are alanine, valine, and threonine; none of these are present in myotoxin *a*.

AMX spin systems were identified by their NH-$C^{\alpha}H$-$2C^{\beta}H$ pattern in the COSY and the lack of further side chain connectivities in the TOCSY. This technique would not preclude misassignment of long chain residues with side chain peaks beyond $C^{\beta}H$ neither visible nor resolved in the spectra. The AMX spin system assignments represented 1 tyrosine, 6 cysteine, 2 histidine, 1 phenylalanine, 3 serine, and 2 aspartate residues. Lacking $C^{\gamma}H$ protons, the aromatic side chains could be connected back to the AMX spin systems by $C^{\beta}H$-ring proton TOCSY (H10, W32, and W34) or NOESY (F12) peaks with the same $C^{\beta}H$ shifts as spin systems identified as AMX. This connectivity was lacking for Y1 and H5.

Glutamate, glutamine, and methionine share an otherwise unique spin system pattern. These 3 residues each appear once in myotoxin *a*. Q3 and

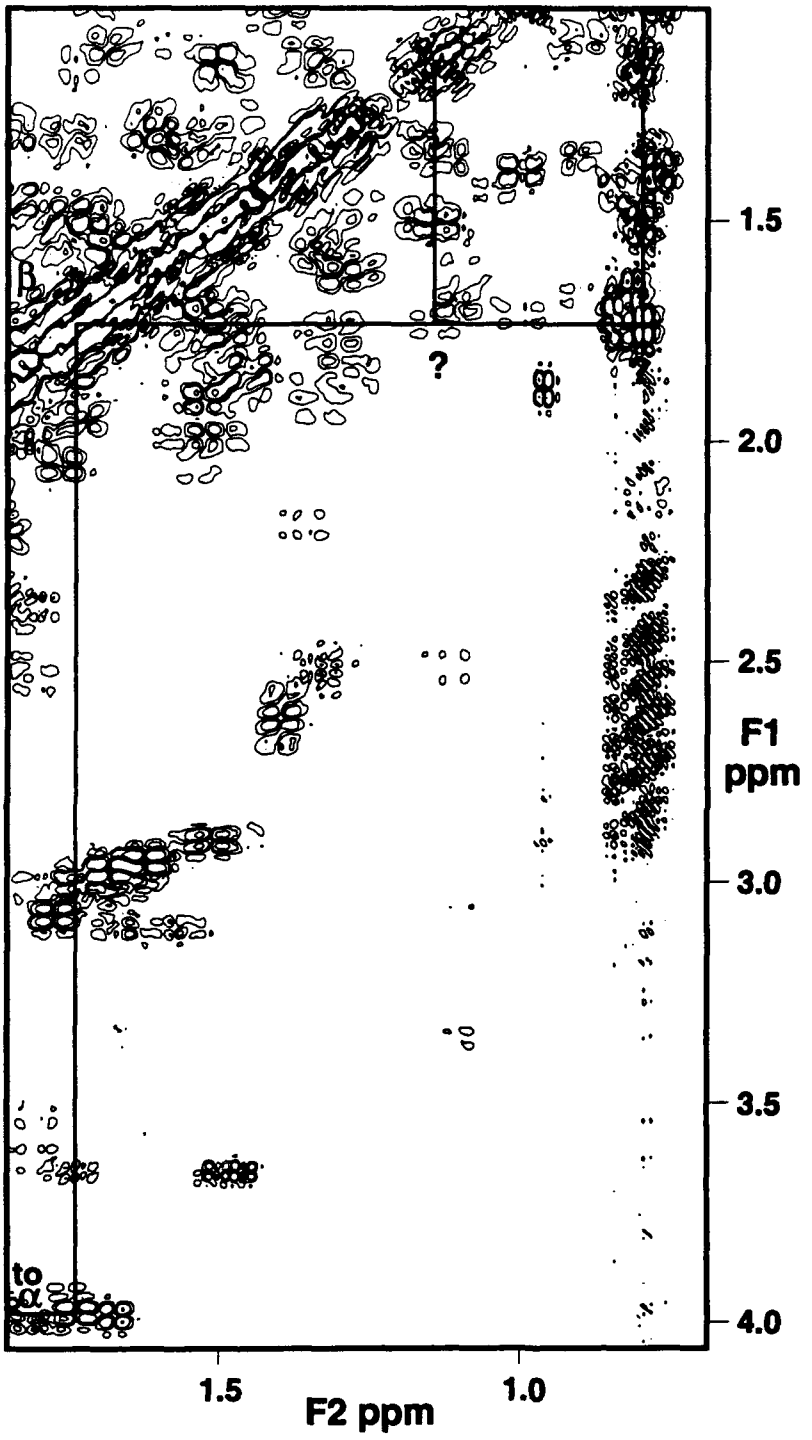


FIGURE 7: I17 side chain connectivities in a DQFCOSY of myotoxin *a*, 3.4mM, pH3.5 in 85% H_2O/15% D_2O at 25°C.

M28 were initially identified as long chain residues and later uniquely identified by their connectivities. The spin system of E15 was identified as E/Q/M in the COSY spectrum by the characteristic pattern (see Figure 8) and later specifically identified by sequential connectivities.

Arginine residues exhibit a $N^{\varepsilon}H$ peak that lies in a sparse area of the spectrum upfield of the aromatic ring proton peaks and downfield of $C^{\alpha}H$ peaks. R31 was uniquely identified by the $N^{\varepsilon}H$-$C^{\delta}H$ peaks which corresponded to $C^{\delta}H$ peaks of the same shift in a long chain spin system.

Serine residues have $C^{\beta}H$ peaks that are shifted unusually far downfield into the $C^{\alpha}H$ region. S22, S23, and S41 were identified as serine residues by being the only 3 AMX residues with $C^{\beta}H$ peaks further downfield than 3.5 ppm (see Figure 6). Sequence specific assignments were made by their sequential connectivities.

A tryptophan residue is readily identified by a $N^{\varepsilon}H$ peak that lies downfield of 10 ppm and forms a COSY crosspeak with the 2h ($C^{\delta 1}H$) peak (nomenclature per Wüthrich, 1986) at ca. 7.2-7.4 ppm. These crosspeaks appear in an uncluttered region of the spectrum and lead into the aromatic region where the NOESY $N^{\varepsilon}H$-7h crosspeak connects to the remainder of the ring proton peaks which can be identified by their unique COSY connectivity pattern. All the ring proton peaks of W32 and W34 were readily identified in this manner (see Figure 9), with their sequence specific assignments made by connectivity back to AMX spin systems and their interresidue connectivities.

The spin systems of histidine residues were identified by three 2h-4h crosspeaks (see Figure 9). One crosspeak was weak compared to the other two and was determined to belong to some minor form of peptide present. Of the two strong peaks, only H10 could be connected back to an AMX spin system and

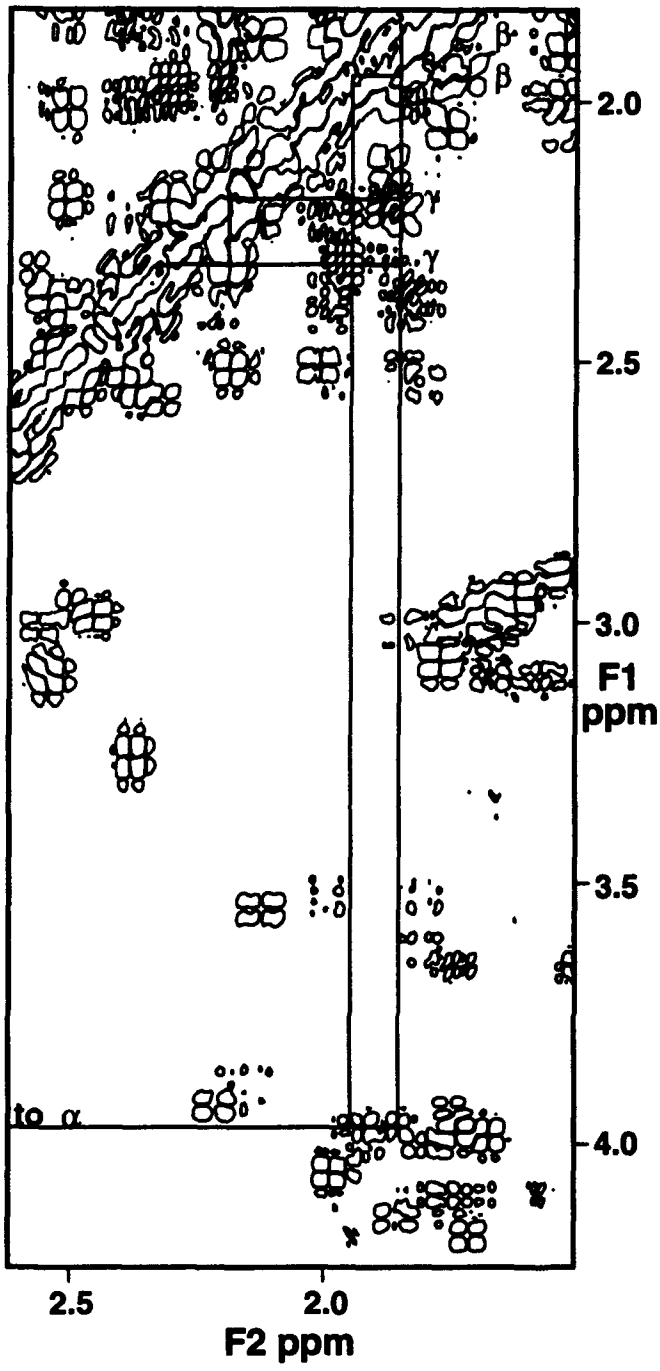


FIGURE 8: E15 side chain connectivities in a DQFCOSY of myotoxin *a*, 3.4mM, pH3.5 in 85% H_2O/15% D_2O at 25°C. Compare to Figure 3.

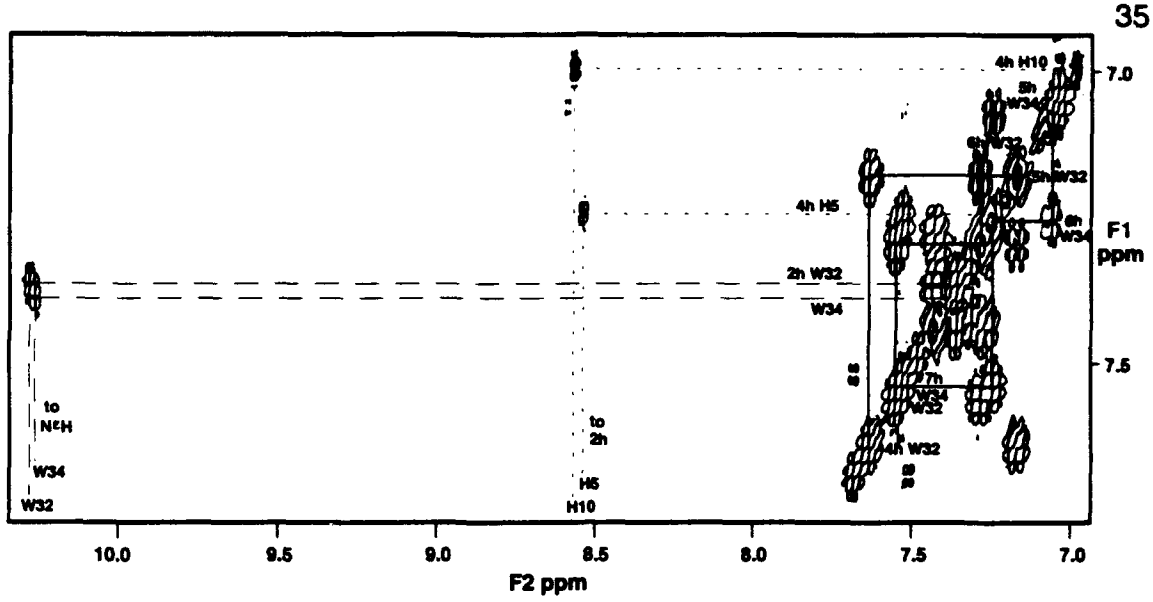


FIGURE 9: Side chain ring proton peaks of W32, W34, H5, and H10 in a DQFCOSY of myotoxin *a*, 3.4mM, pH3.5 in 85% H_2O/15% D_2O at 25°C.

its sequential connectivities. H5 was, therefore, assigned by default. Consequently, the assignment of AMX peaks for H5 are, admittedly, less sure.

The aromatic spin system of F12 was readily identified by its characteristic COSY pattern (see Figure 10). The corresponding AMX spin system was initially identified by sequential connectivities and later corroborated by a $C^{\beta}H$-2,6h NOESY crosspeak.

The spin systems of the three proline residues were identified after all other spin systems were brothered. Since proline lacks an amide proton, the characteristic $C^{\alpha}H$-$C^{\beta}H$-$C^{\gamma}H$-$C^{\delta}H$ COSY connectivities, with $C^{\delta}H$ shifts lying between $C^{\alpha}H$ and $C^{\gamma}H$ shifts, were identified in the upfield half of the spectrum (see Figure 11) without corresponding connectivities to a NH peak. The specific identification of P13 and P21 resulted from sequential $C^{\alpha}H_i$-NH_{i+1} NOESY crosspeaks to K14 and S22, respectively. P20 was assigned as the remaining proline spin system.

As mentioned earlier, the fingerprint NH-$C^{\alpha}H$ peaks for D29 and K35 were visible in only the 50°C spectra (see Figure 12) because of their proximity to the water resonance at 25°C.

Wüthrich (1986) predicted that the connectivities soley in the fingerprint region of a NOESY spectrum would result in correct sequential assignments for about one half of the residues. For this reason and to resolve ambiguities, a search for sequential connectivities was carried out in the NH-NH and NH-$C^{\beta}H$ regions of the NOESY spectrum. Figure 13 shows several such NH-NH connectivities.

The spin system, aside from the ring, was not identified for Y1, due to the highly exchangeable amino protons and also perhaps due to a highly flexible N-terminus. The assignments of residues K2, Q3, H5, and K6 are less certain than

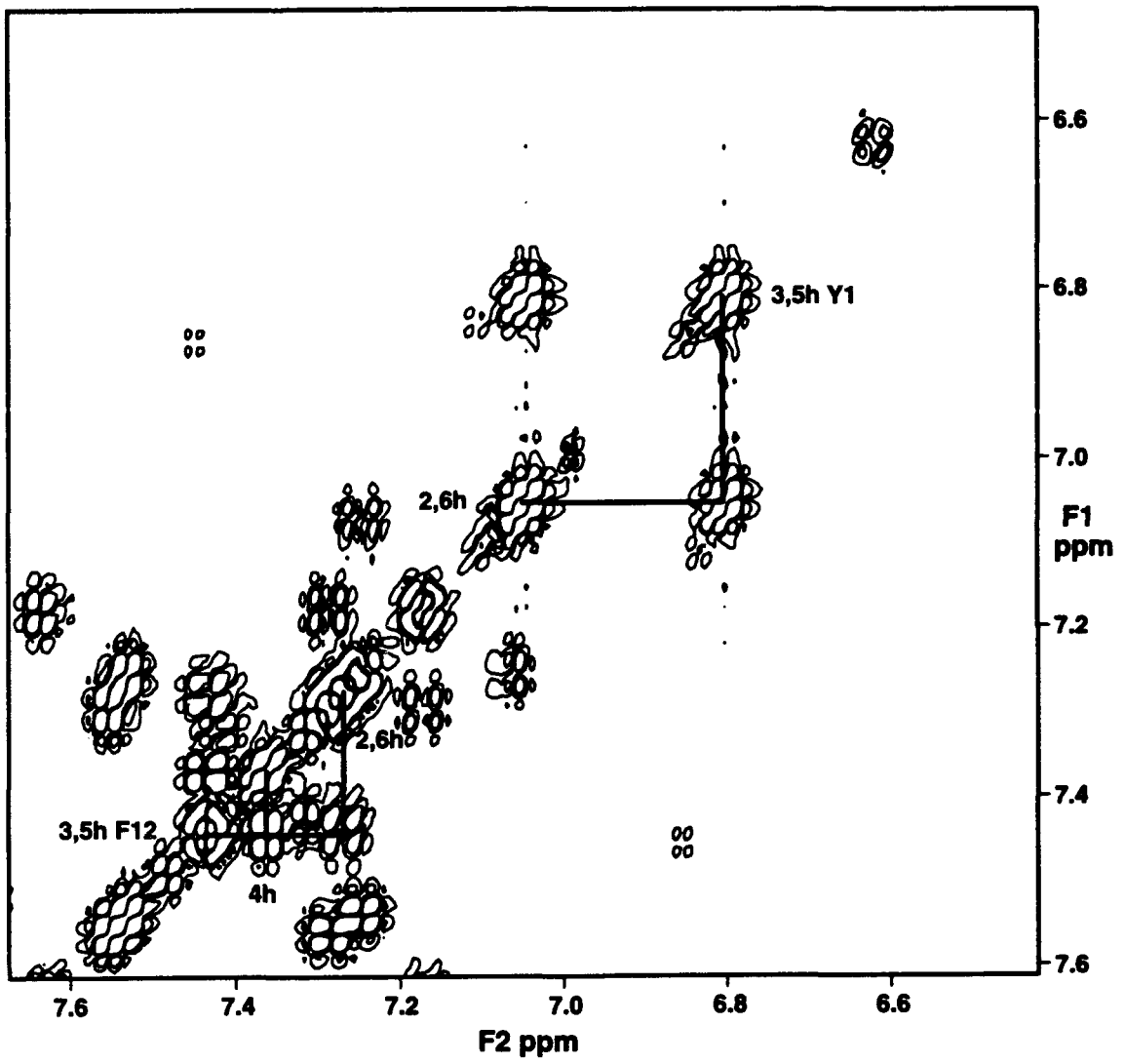


FIGURE 10: Aromatic ring proton peaks of F12 and Y1 in a DQFCOSY of myotoxin *a*, 3.4mM, pH3.5 in 85% H_2O/15% D_2O at 25°C.

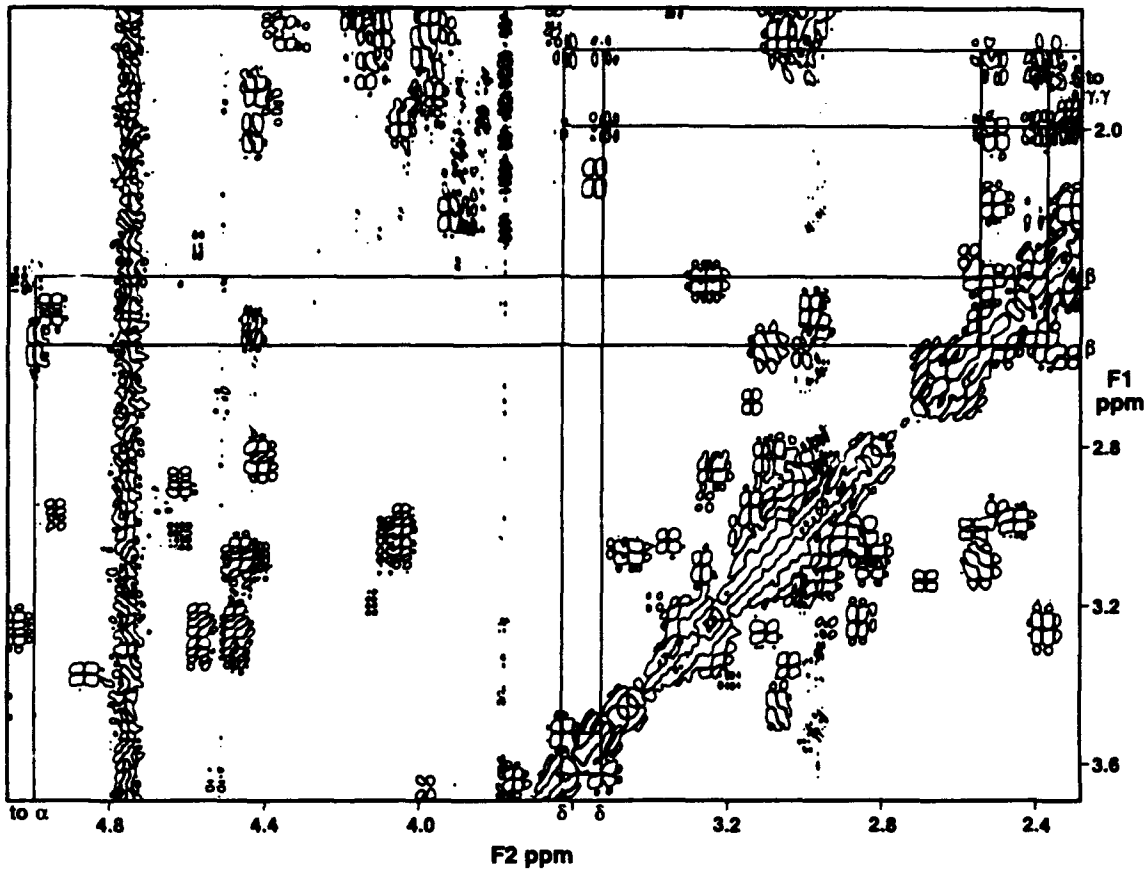


FIGURE 11: P20 side chain connectivities in a DQFCOSY of myotoxin *a*, 3.4mM, pH3.5 in 85% H_2O/15% D_2O at 25°C.

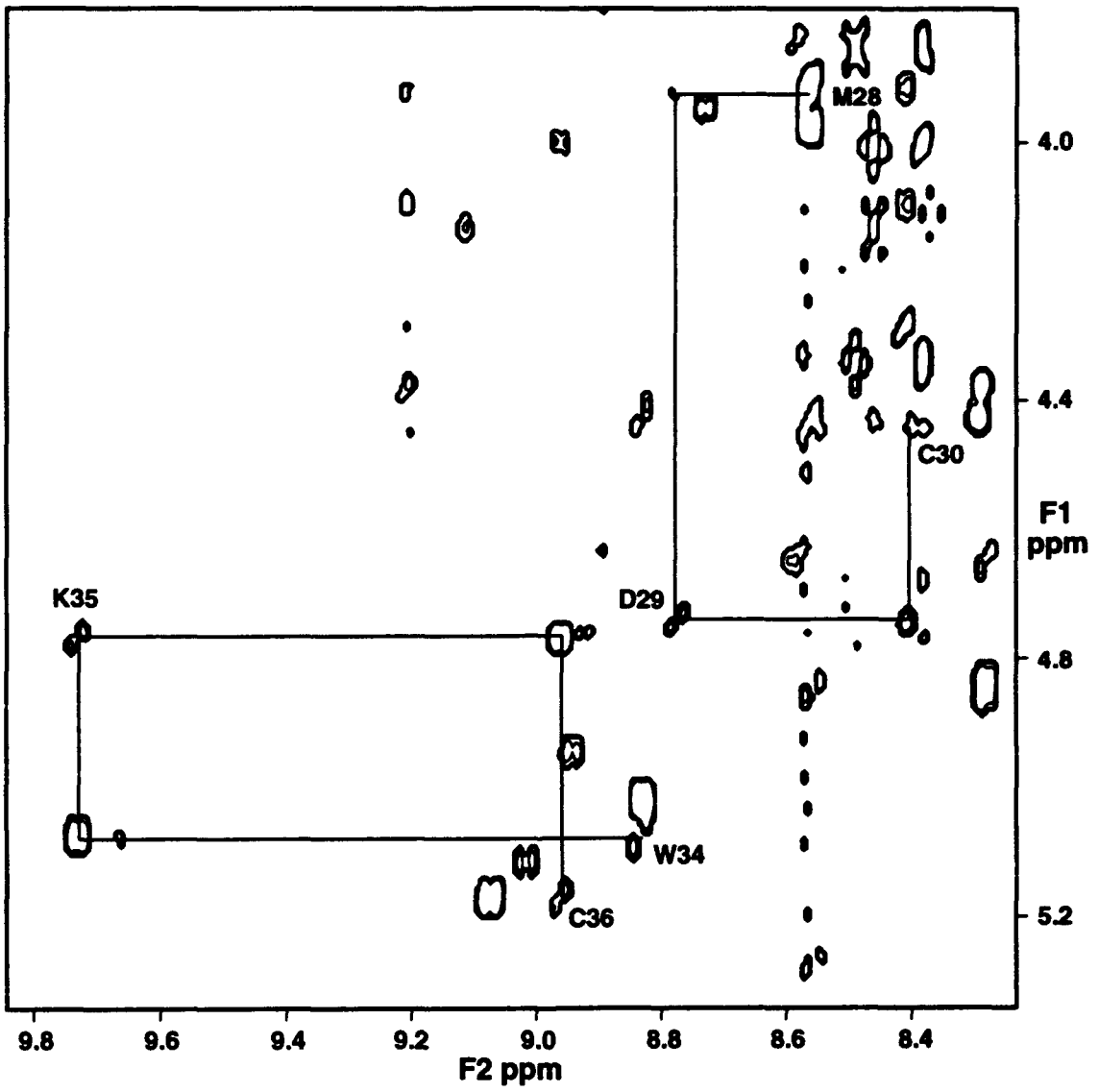


FIGURE 12: Intraresidue and sequential NH-C$^{\alpha}$H peaks for D29 and K35 in NOESY spectrum of myotoxin *a*, 3.7mM, pH 3.5 in 85% H_2O/15% D_2O at 50°C.

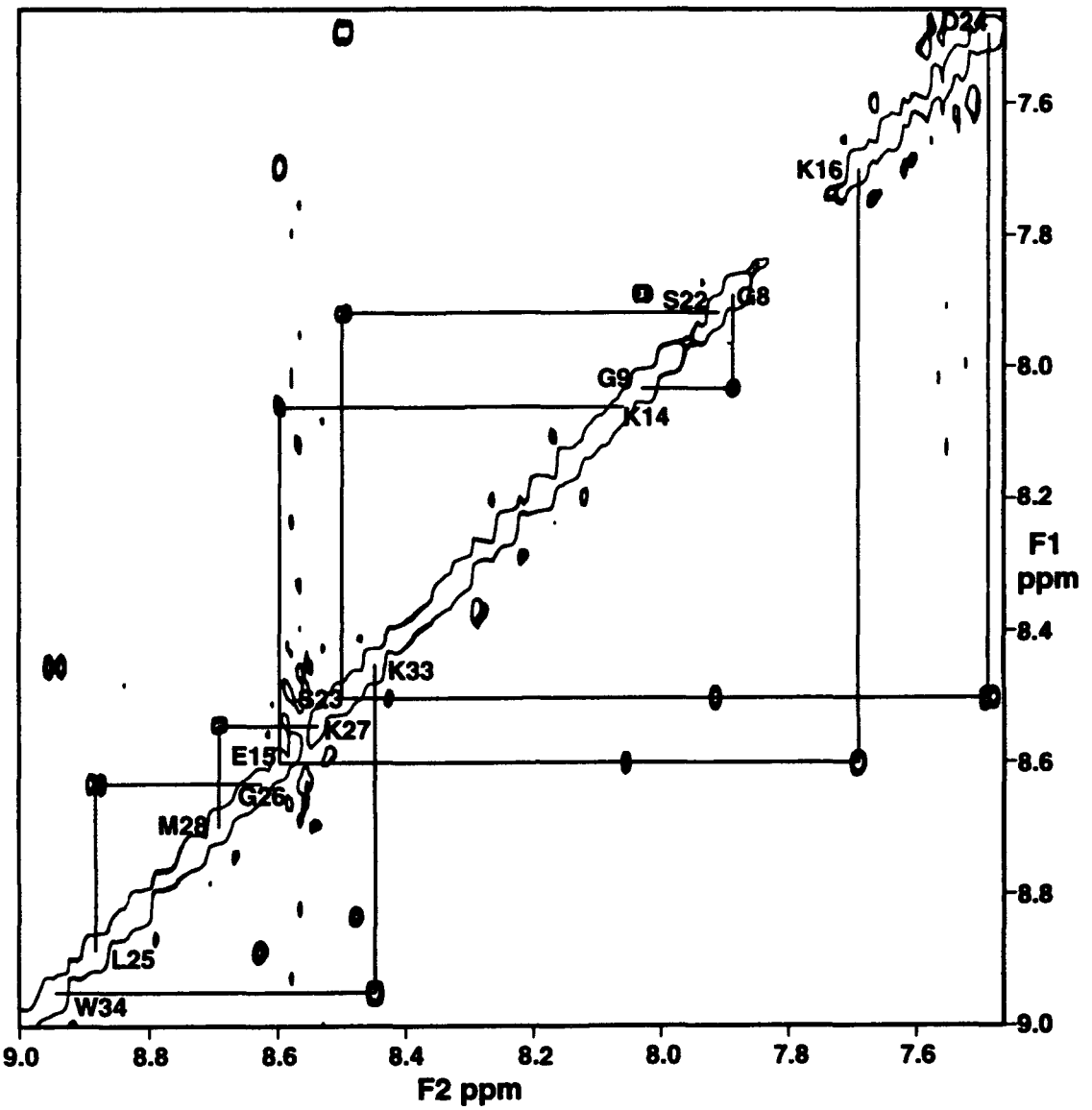


FIGURE 13: NH-NH region of NOESY (200ms) of myotoxin *a*, 3.4mM, pH 3.5 in 85% H_2O/15% D_2O at 25°C. Sequential NH-NH connectivities are shown.

the rest of the assignments because they were determined solely by fingerprint region connectivities and spin system (long chain and AMX). G42 was not identified since connectivity could not be established from S41 to any of the remaining 3 pairs of glycine peaks. The remaining assignments were corroborated with additional connectivities or identification of unique spin systems or both. The assignment of the fingerprint region of the COSY spectrum appears in Figure 14. Many unassigned, excess, unknown peaks are clearly evident. The sequential connectivities for K7-F12, P13-I19, and P21-S41 are shown in Figure 15. The resultant sequence specific ^{1}H-NMR spin system assignments are summarized in Table 1.

Distance Geometry. The original constraints (Nieman, unpub.) contained a few highly questionable assignments. These few constraints were removed, and Dspace structures were generated. This process was continued with modifications to the Dspace refinement macros and minor changes to the constraints, such as constraining only one of pair of diastereotopic protons when only one such volume was measured. Eventually, the best structures from Dspace were created with this file of 170 constraints (44 intraresidue, 74 interresidue sequential, and 52 interresidue nonsequential). Ten pairs of structures were created, each structure taking ca. 15 hours to complete. The four most well refined structures superimposed on the best structure gave all atom root mean square distances (RMSDs) of 4.71, 4.52, and 5.02 Å. The backbones of these superimposed structures are visualized in Figure 16.

Since an alternate disulfide bond arrangement had been published for the highly homologous myotoxin, crotamine (Conti & Laure, 1988), structures were created in Dspace with the same experimental constraints but without disulfide bonds. Arbitrary disulfide bonds were selected from the cysteine side chains in

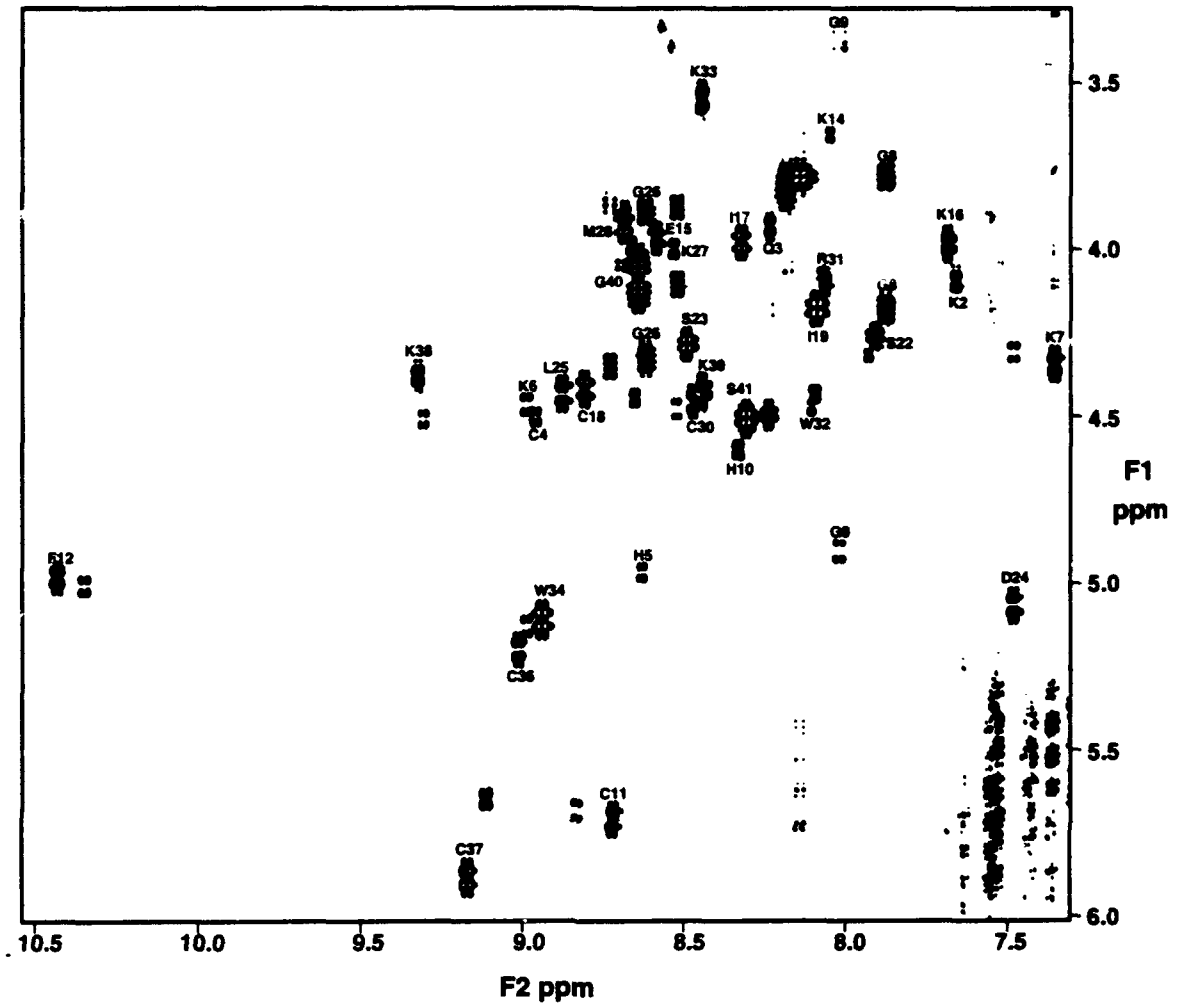


FIGURE 14: Fingerprint region (NH-C$^{\alpha}$H) of DQFCOSY of myotoxin *a*, 3.4mM, pH 3.5 in 85% H_2O/15% D_2O at 25°C. Unlabeled peaks are of unknown origin.

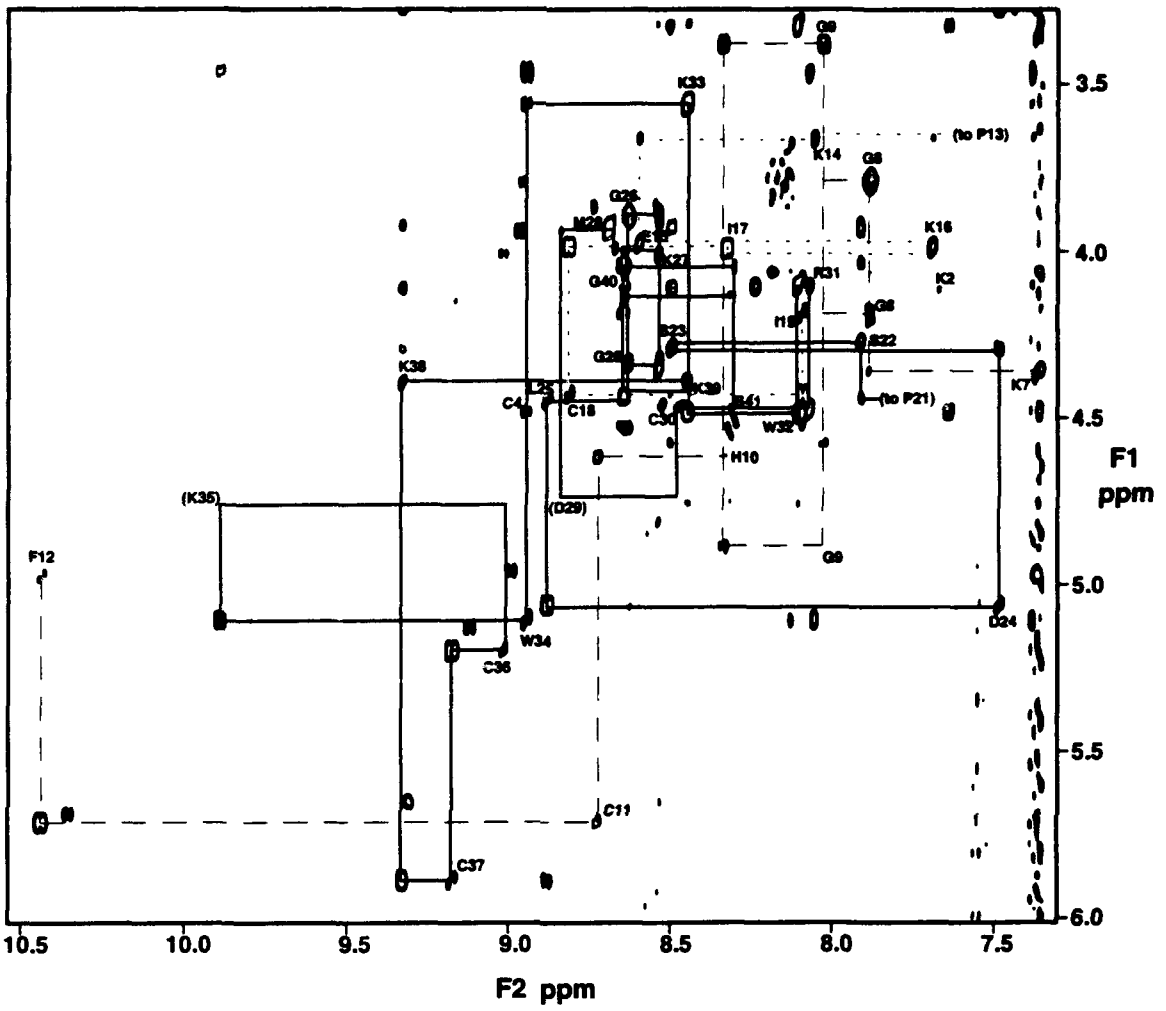


FIGURE 15: Fingerprint region (NH-C$^{\alpha}$H) of NOESY of myotoxin *a*, 3.4mM, pH 3.5 in 85% H_2O/15% D_2O at 25°C. Intraresidue NH-C$^{\alpha}$H peaks are labeled; sequential interresidue NH-C$^{\alpha}$H peaks are not labeled. Residues whose labels appear in parentheses are visible in the 50°C spectra.

Table I: ^{1}H NMR Chemical Shifts of Myotoxin *a* at pH 3.5, 25°C (*50°C) (ppm).

residue	NH	$C^{\alpha}H$		$C^{\beta}H$		$C^{\gamma}H$		$C^{\delta}H$		other (and unassigned)					
Tyr-1										7.05(2,6h)	6.80(3,5h)				
Lys-2	7.67	4.10								1.63	1.15				
Gln-3	8.24	3.94								1.75	0.85				
Cys-4	8.97	4.52		3.16	2.69										
His-5	8.63	4.96		2.98	2.51										
Lys-6	8.99	4.48		1.39						1.88					
Lys-7	7.36	4.35		1.76						2.06	1.54				
Gly-8	7.89	4.19	3.78												
Gly-9	8.03	4.88	3.38												
His-10	8.34	4.62		3.03	2.89					8.58(2h)	6.99(4h)				
Cys-11	8.73	5.71		3.14	2.96										
Phe-12	10.44	4.97		2.99	2.46					7.44(3,5h)	7.37(4h)	7.28(2,6h)			
Pro-13		3.66		1.75	1.50	1.32	1.16	2.91							
Lys-14	8.05	3.66													
Glu-15	8.60	3.97		1.97	1.88	2.31	2.19								
Lys-16	7.69	3.98		1.67						1.36	1.11	0.91			
Ile-17	8.33	3.98		1.74						1.15	0.80				
Cys-18	8.82	4.43		3.11	2.83										
Ile-19	8.09	4.18		1.71		0.85	0.77								
Pro-20		5.01		2.54	2.37	1.99	1.80	3.63	3.52						
Pro-21		4.44		2.51	2.01	2.23	2.14	3.87							
Ser-22	7.91	4.27		4.03	3.92										
Ser-23	8.50	4.29		4.11	3.92										
Asp-24	7.49	5.06		3.26	2.39										
Leu-25	8.88	4.45		1.84	1.45					1.75	0.84				
Gly-26	8.63	4.33	3.89												
Lys-27	8.54	4.00		1.92	1.79	1.55	1.34								
Met-28	8.69	3.93		2.23						2.51	2.40	1.86			
Asp-29	8.77*	4.74*													
Cys-30	8.48	4.47		3.15	3.08										
Arg-31	8.07	4.11		1.57	1.32			2.55	2.49	6.62(nε)					
Trp-32	8.11	4.48		3.32	3.25					10.28(nε)	7.64(4h)	7.56(7h)	7.37(2h)	7.30(6h)	7.18(5h)
Lys-33	8.45	3.55		2.13						1.41	0.79	0.58			
Trp-34	8.94	5.11		3.47	3.07					10.27(nε)		7.54(7h)	7.39(2h)	7.26(6h)	7.07(5h)
Lys-35	9.89	4.76*								1.62	1.27				
Cys-36	9.02	5.20		3.24	2.86										
Cys-37	9.17	5.88		3.27	3.11										
Lys-38	9.33	4.39		1.95	1.75					1.54	1.46				
Lys-39	8.45	4.43		1.91						3.01	1.72	1.53			
Gly-40	8.64	4.14	4.04												
Ser-41	8.31	4.52		3.91	3.83										
Gly-42															

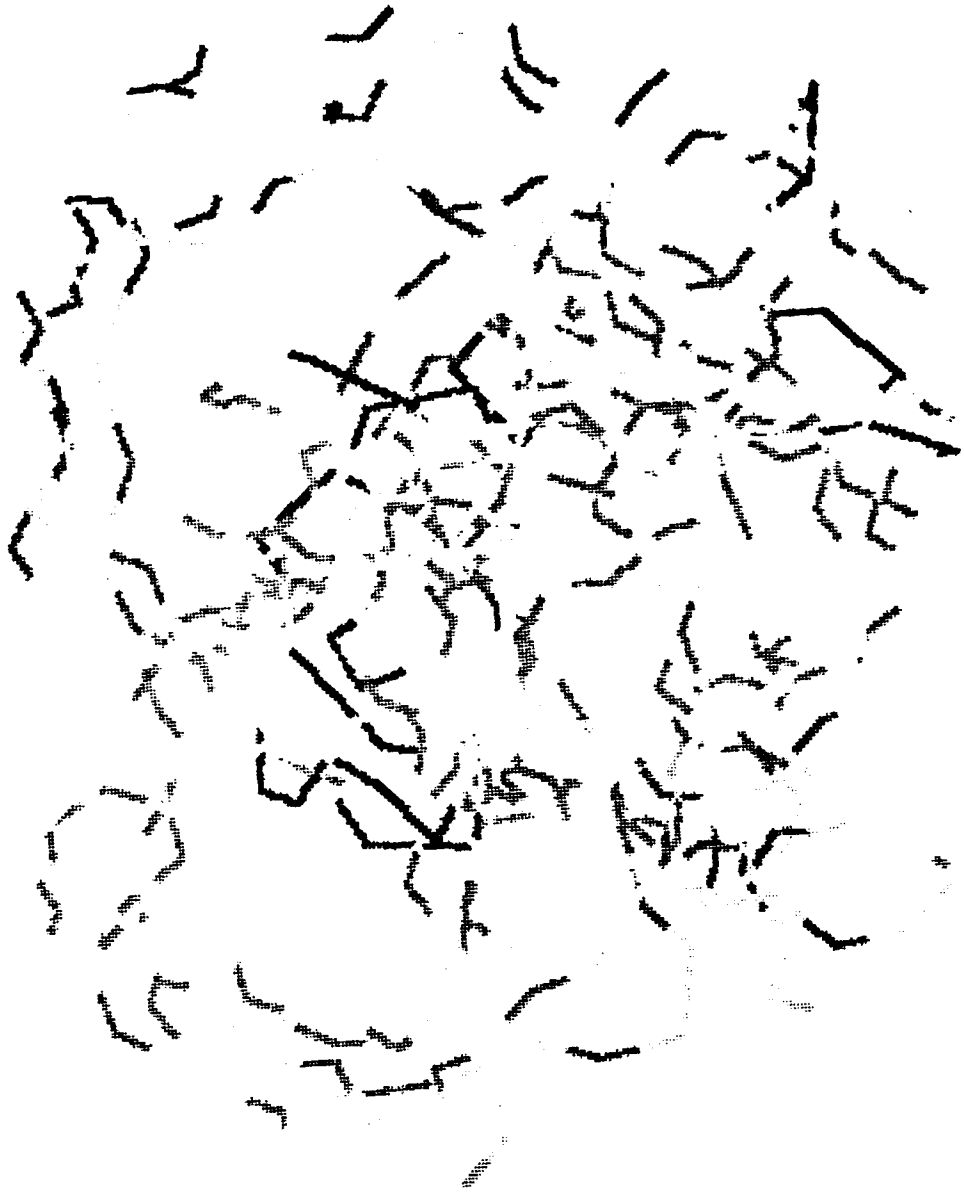


FIGURE 16: Superimposition of four of the best Dspace structures (backbones only) of myotoxin *a* based on the original distance constraints (see text).

closest proximity in the resultant best structure. More Dspace structures were generated with the same experimental constraints but with the new disulfide arrangement of 4-18, 11-36, and 30-37. Surprisingly, these structures were both better refined and had lower RMSDs. The best four superimposed on the best with all atom RMSDs of 3.55, 3.34, and 3.25 Å. This result led to a reinvestigation of the disulfide bond arrangement (Bieber et al., unpub.) which confirmed the original structure of Fox et al. (1979). None of the Dspace structures from this set of original constraints, whether with the actual or the arbitrary disulfide arrangement, showed much secondary structure. They looked like three loops constrained by these disulfide bridges.

A possible reason for the lack of converged structures could have been incorrect sequential assignments. The sequential assignments described in the previous section were independently made with new spectra and compared with the assignments of previous investigators (Henderson, 1986; Murchison, 1989; Nieman, unpub.). The results were in near complete agreement with previous results; the major exception was the uncertainty in the assignment of the N-terminal residues mentioned in the previous section.

DIANA proved to be a much more useful distance geometry program because it ran much faster, filtered out meaningless constraints, and provided better feedback which allowed one to locate potentially erroneous constraints. The original constraints, previously used with Dspace, were put in DIANA format and run as upper limits. DIANA put out a modified version of this upper limit file after it had deleted irrelevant constraints, lengthened constraints that were shorter than those allow by the covalent geometry, and lengthened constraints to accommodate the use of pseudoatoms for non-stereospecifically assigned prochiral protons. The resulting file consisted of 116 experimental constraints

(eight intraresidue, 59 interresidue sequential, and 49 interresidue nonsequential). Running on the Cray, 10 structures were created in ca. 11 min, consuming 2.33 min of CPU time. The lowest target function value was 69.65, and the best pair of superimposed backbones had a RMSD of 4.10Å.

Several iterations of removing consistently violated constraints and building structures led to a final version of the original constraints, containing 88 experimental upper limits (8 intraresidue, 52 interresidue sequential, and 28 interresidue nonsequential) (see Appendix C). None of these experimental constraints involved Y1, K2, Q3, H5, K6, or K7. From these constraints, 536 structures were created, each taking an average of just under 3 min clock time or 1.06 min CPU time, on the Indigo. The best structure had a target function of 0.71, or almost 2 orders of magnitude better than the first version from original constraints. This result seemed to indicate that many erroneous constraints had been weeded out. The 4 best structures superimposed on the best structure with backbone RMSDs of 2.25, 2.80, and 2.90 Å.

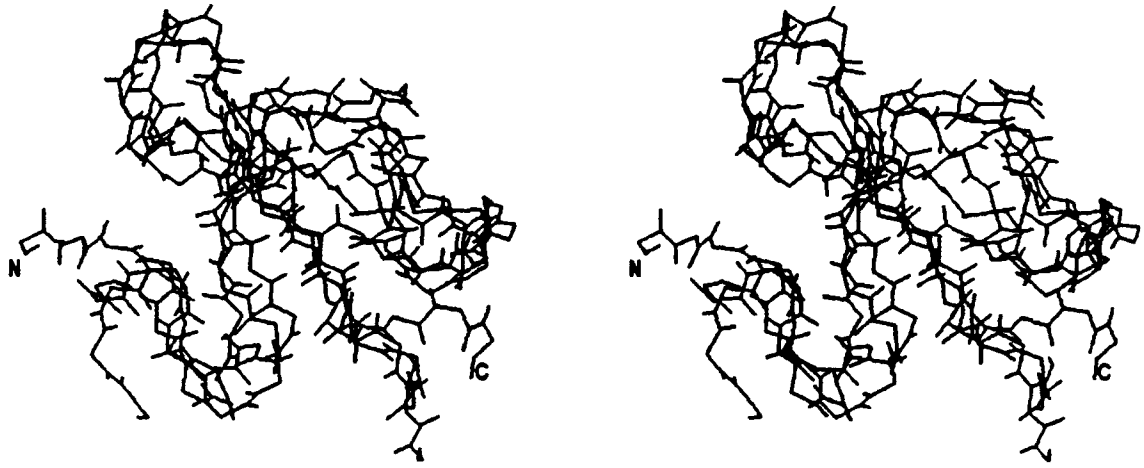
Using the confirmed assignments, 200 ms NOESY peaks were integrated and converted to distance constraints in Felix. These constraints were manually converted to a DIANA upper limit file of 134 experimental constraints (77 intraresidue, 29 interresidue sequential, and 28 interresidue nonsequential). Several additional interresidue nonsequential constraints would have been attained if NHK7, 2hW32, and 4hF12 had not had so many ambiguous crosspeaks. This set of new constraints shared only 36 constraints (6 intraresidue, 14 interresidue sequential, and 16 interresidue nonsequential) with the original set, all of which were modified. When run with the new constraints, DIANA put out a modified version that deleted 40 meaningless constraints and modified 49 for non-stereospecific assignments. The resultant set of new

constraints consisted of 86 experimental upper limits (31 intraresidue, 24 interresidue sequential, and 31 interresidue non-sequential) (see Appendix C).

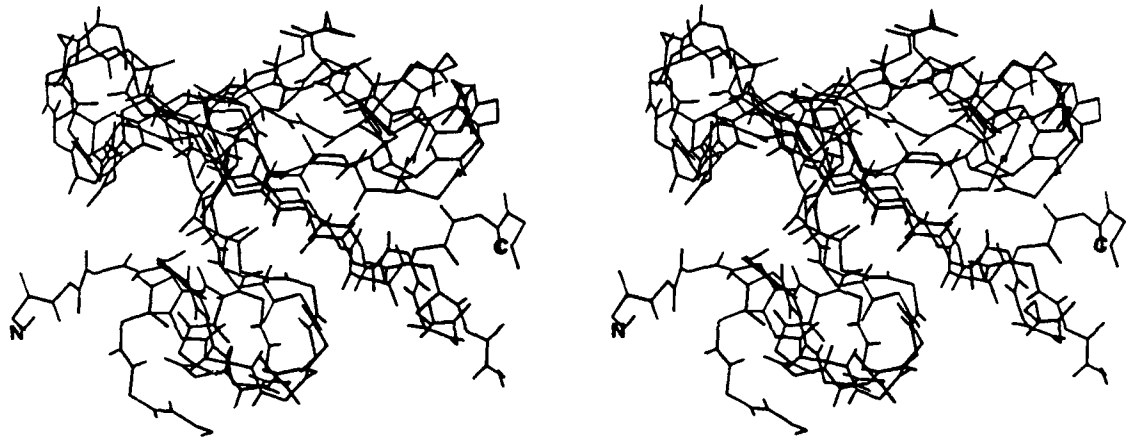
These new constraints were run in DIANA to produce 999 structures. The lowest target function was 0.49. The 4 best structures superimposed on the best structure with backbone RMSDs of 2.52, 2.30, and 2.41 Å (all non-H atom RMSDs of 4.10, 3.79, and 3.87 Å). When the best structure from the original constraints and the second and third best structures from the new constraints are superimposed, they show a consistent backbone folding with RMSDs of ca. 2.5 Å (see Figure 17a), in spite of sharing only 42% of the same constraints. Figure 17b shows the best structure from the original constraints and the best and third best structures from the new constraints superimposed. In both cases, a central core of three foldings of antiparallel β-sheet (see Figure 18) constrained by 3 disulfide bridges are joined by loops and turns, including a modified type VI turn (Creighton, 1993) in residues C18, I19, P20, and P21 (see Figure 19). This unusual turn was identified by looking for and finding a strong NOESY crosspeak between C^{α}HI19-C^{α}HP20, uniquely indicative of a *cis*-proline peptide bond (see Rawn, 1989, for stereoviews of *cis* and *trans* proline peptide bonds). Neither of the other two prolines showed such a peak. It is possible that a *cis-trans* isomerization about this bond could be the source of two conformations observed as interconverting peaks on the HPLC.

Low Temperature HPLC Separations. Separation of 1.0 mg of myotoxin *a* (10 mg/ml in water) by semipreparative RP-HPLC at 25°C on a 20-21% acetonitrile gradient yielded two large peaks (A and B) in a 4:1 ratio

$\frac{B}{A+B} = 80.4\%$ (see Figure 20a). After drying, redissolving, and injecting the



a



b

FIGURE 17: Stereoviews of three superimposed backbones of distance geometry generated structures of myotoxin *a*. See text for details.

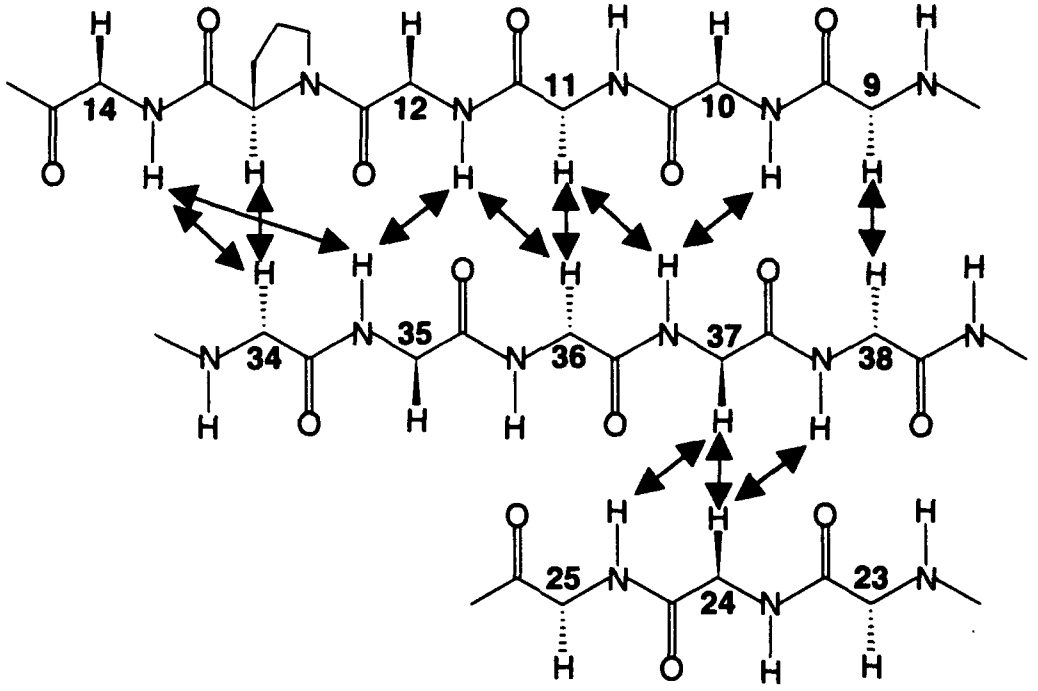


FIGURE 18: Schematic representation of antiparallel β-sheet in myotoxin *a* with arrows pointing to proton pairs which give rise to NOE peaks.

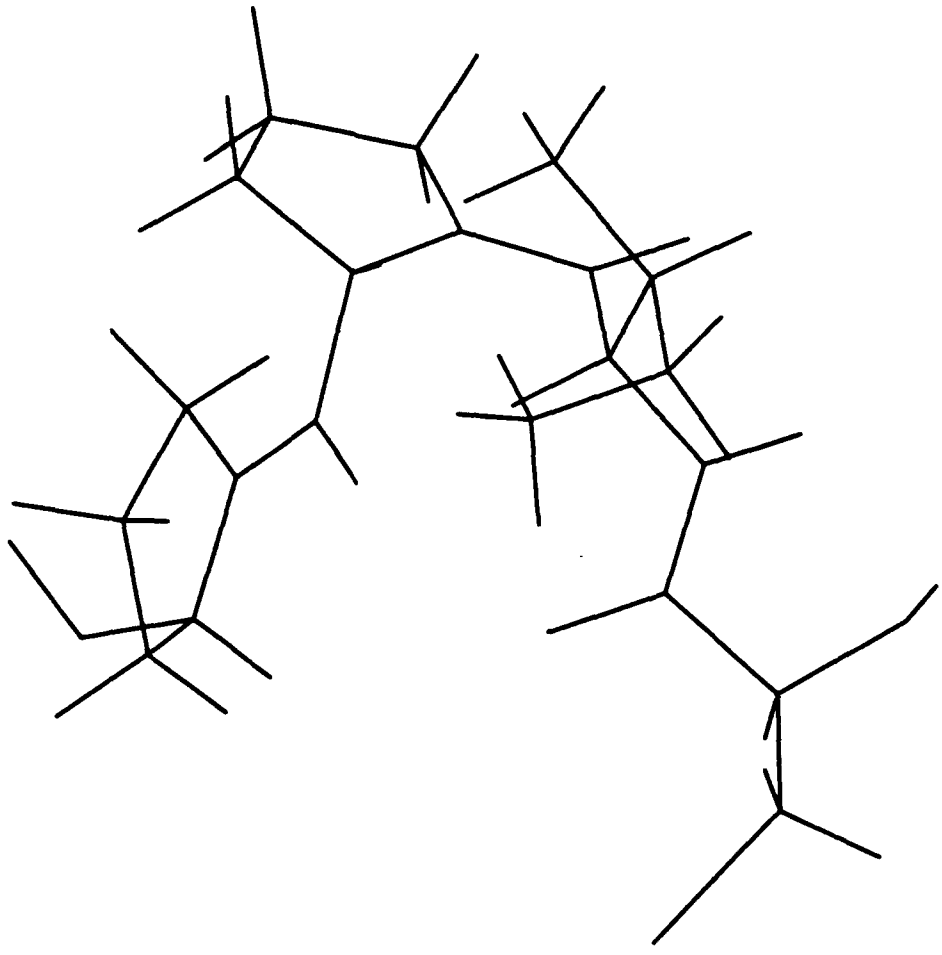


FIGURE 19: Modified type VI turn along C18-I19-P20-P21 with a *cis*-peptide bond between I19-P20. Because of P21, the turn lacks normal hydrogen bonding.

B fraction of this and product from two similar runs (77.3% and 83.1% B), separation yielded 79.4% B (Figure 2b). 500 µl of this B fraction was injected 71 min later than the previous run. This separation yielded 83.4% B (Figure 20c), indicating that the B fraction equilibrates back into the A and B fractions within about an hour at 25°C.

When using chilled solvents and sample, separation of 1.0 mg of myotoxin *a* (10 mg/ml in water) at ca. 2°C on a 24-27% acetonitrile gradient yielded 70.7% B (Figure 20d). When this B fraction was held on ice and 500 µl of it was injected 206 min later than the previous run, the separation yielded 95.7% B (Figure 20e). This result indicates that at 2°C the rate of reestablishing equilibrium is significantly reduced, maintaining an enrichment of over 95% B for more than 3.4 hours.

GCG Analyses. Peptide sequence homology searching using FastA in the SwissProt database revealed five proteins with >80% homology. These were all rattlesnake myotoxins. A gap existed in the homology scoring from >80 to 60, confirming the uniqueness of this family (Bieber et al., 1987). The other homologies were over small regions, usually either in the first or second half of myotoxin *a*'s sequence but rarely in the middle (see Appendix D). Of functional interest, a 60% homology over 10 residues, including 3 cysteines, with additional conservative substitutions exists with rat brain sodium channel protein II (see Figure 21). A TFastA search of the GenEMBL database revealed that transcriptions of high homology or obvious functional relevance to myotoxin *a* were not present. A search of Motifs identified myotoxins and a cAMP/cGMP-dependent protein kinase phosphorylation site.

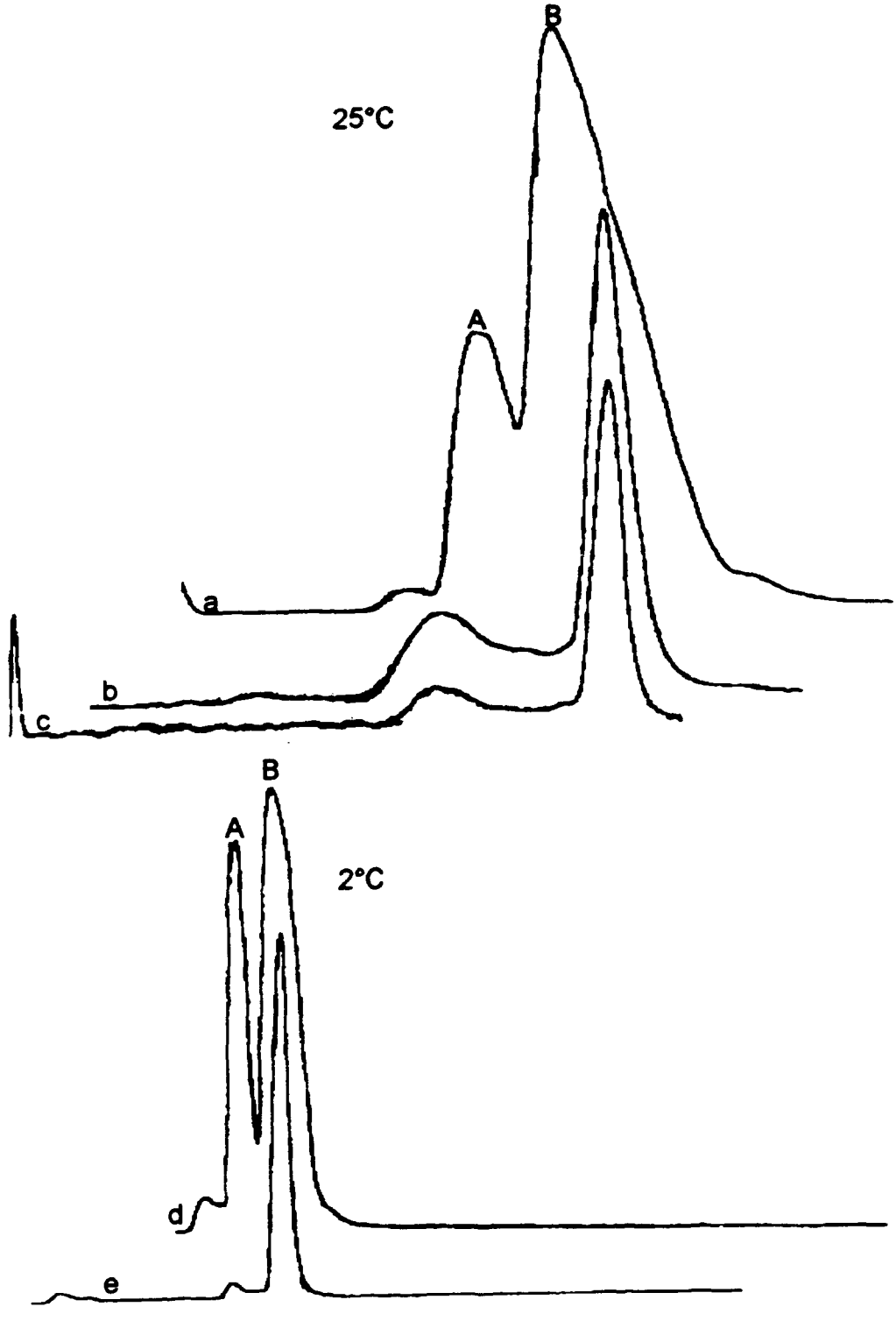


FIGURE 20: RP-HPLC spectra of myotoxin *a* at (A) 25°C and (B) 2°C. See text for details.

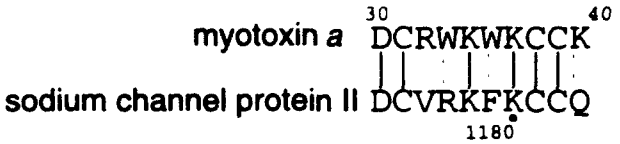
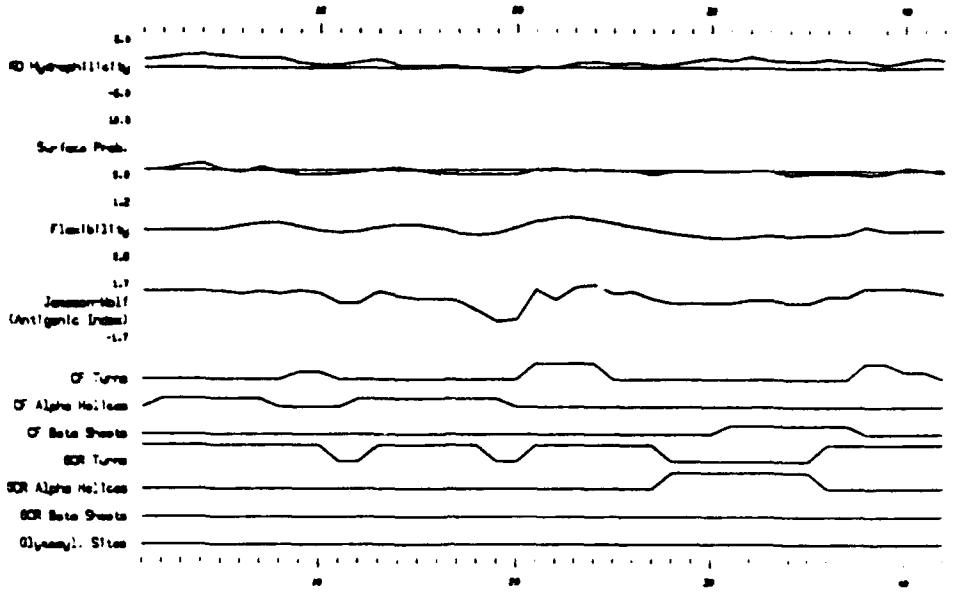


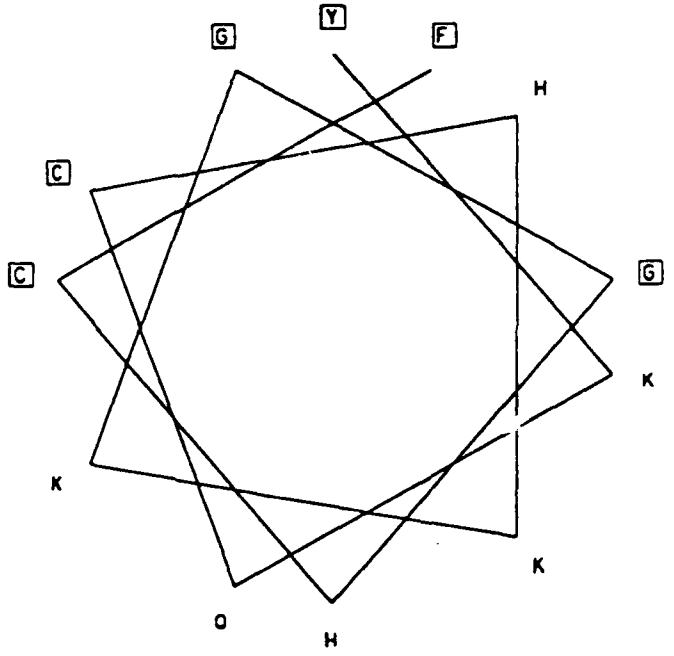
FIGURE 21: Region of homology between myotoxin *a* and rat brain sodium channel protein II. Dashed lines indicate conservative substitution per FastA.

PeptideStructure provided a Chou-Fasman prediction of α-helix (weak helical formers) for residues 2-7 and 12-19; turns at 9-10, 21-24, and 38-41; and β-sheet (weak β-sheet formers) at 31-37. This prediction is in contrast with the Garnier-Osguthorpe-Robson prediction of nearly all turns except for a helical stretch from 28-35 (see Figure 22a). HelicalWheel was performed on the possible helical region1-12, taking into account the predictions, locations of prolines, and the distance geometry structures. Figure 22b shows that if this region were an α-helix, a hydrophobic "greasy" patch would exist on one side of the helix while many charged residues would be on the opposite side.



a

□ = Hydrophobic



b

FIGURE 22: GCG PlotStructure secondary structure predictions for myotoxin *a* (a, top) and HelicalWheel (b, bottom) for residues 1-12.

CHAPTER IV

CONCLUSIONS

Structural Features. The distance geometry generated structure of myotoxin *a* is not highly defined. The most well refined structures superimpose with backbone RMSDs of ca. 2.5 Å, a rather course "resolution," especially for such a small protein. Nevertheless, the fact that these structures were generated from two sets of distance constraints which were derived from different spectra, separately assigned, and only 42% alike leads to a high degree of confidence in the accuracy of the global folding. The confirmed sequence specific assignments likewise seem fairly certain.

The lack of tighter convergence of structures comes from two sources. First, the distance constraints used are a first approximation from NOESY data at a single mixing time. More accurate first approximations could be made by assigning a set of DQCOSY, TOCSY, and NOESY spectra taken on the same sample under the same conditions, with several mixing times for NOESY spectra. Using the tools of Felix, the actual integration and subsequent derivation of distance constraints from NOESY spectra that differ only in mixing times is a straightforward procedure.

Even with multiple mixing times, an inherent inaccuracy exists in deriving distance constraints using this isolated spin pair approximation, where the intensity of a NOE peak is assumed as inversely proportional to the sixth power of the distance between two interacting protons. This approximation fails to account for spin diffusion, the transference of magnetism to other nearby protons. This effect can significantly alter the distance constraints derived from an isolated spin pair approximation for mixing times as short as 50 ms (Meadows et al., 1991).

One readily available way to refine the distance constraints to a greater level of accuracy is through the use of BKCALC (Hare Research) (Nerdal et al.,

1989), a module within Dspace. BKCALC takes a structure generated from a set of first approximation distance constraints and back calculates its NOESY spectrum, taking into account spin diffusion. The calculated spectrum is compared with the experimental spectrum, and distance constraints are adjusted accordingly. More structures are generated from these refined constraints, and the process is repeated in an iterative manner until the calculated spectrum matches the experimental spectrum.

The second source of limited convergence of structures is the relatively few nonsequential distance constraints, those which are most important in defining the overall folding of the protein (Wüthrich, 1986). It simply would not be possible to highly define the positions of all side chains in a protein this size with 86 total constraints, 31 of which are nonsequential. As mentioned earlier, a few ambiguities prevented the complete assignment and, hence, use of several partially assigned NOE peaks and their resultant distance constraints. As is apparent at this point, the excess peaks introduced by the chemical microheterogeneity and isomerization have their most detrimental effect as distracters causing ambiguities in the NOESY spectra. Distance constraints should be used from only unambiguously assigned peaks, as use of erroneous constraints would be highly counterproductive to distance geometry calculations and waste much time.

Chemical Microheterogeneity. The slight differences between the forms of *C. v. viridis* myotoxins make it difficult to separate them by conventional preparative biochemical techniques, such as ion exchange chromatography, RP-HPLC, gel filtration, or precipitation techniques. Clearly, the greatest hope for purification of myotoxin *a* from the other *viridis* forms lies in affinity chromatography. A synthesized C-terminal fragment corresponding to *viridis*-3

could be attached to a carrier protein and used for the production of polyclonal antibodies specific to *viridis*-2, 3, and 4 but not to myotoxin *a*. It is not certain whether a fragment small enough to elicit such a specific response would elicit much of a response in the first place. Conversely, the entire myotoxin *a* sequence could be synthesized, but the likely success of properly joining the cysteines and folding the protein is also uncertain.

Isomers. The presence of two isomers (as observed by HPLC) which interconvert slowly on the NMR timescale leads to a spectrum that is the population weighted sum of the resonances from the two isoforms, rather than their average. The greatly reduced rate of interconversion at 2°C offers hope of keeping a sample highly enriched in one form long enough to acquire NMR spectra (1D or, preferably, 2D). The exploration of solvent systems which might greatly shift the equilibrium to one conformer or the other seems worthy of consideration as well.

The observation of a *cis*-proline peptide bond leads to consideration of *cis-trans* isomerization as a prime suspect for producing the two observed conformers. The use of peptidyl proline isomerase ((PPI) (Harrison & Stein, 1990) in kinetic experiments that employ HPLC and 1D ^{1}H-NMR methods (Hsu et al., 1990) to measure the rates of interconversion would help test this hypothesis. An understanding of the interconversion mechanism, however, does not seem directly relevant to solving the structure.

It is not possible to determine the extent to which chemical microheterogeneity and isomerization each affect the spectra. Therefore, it is not possible to predict how much the ^{1}H-NMR spectra would improve by producing a chemically homogeneous sample or by maintaining a sample enriched in one conformer. Certainly, combining both approaches should lead to

^{1}H-NMR spectra dominated by a single species and allow a fairly quick analysis of the spectra and solution of a highly defined structure. The pursuit of both approaches would be very time and resource consuming. It is possible that a single approach may remove many plaguing peaks from ^{1}H-NMR spectra, leading to more complete assignments and enough distance constraints to build well converged and refined structures. The most pragmatic approach would be to try the easiest strategy first: low temperature isomer enhancement by RP-HPLC. The use of three-dimensional ^{1}H-NMR experiments may also prove to be a relatively easy way to resolve spectral ambiguities.

Structure-Function. The combination of solved structural features and secondary structure predictive methods along with the pH titration work of Henderson (1986) leaves open the possibility of a N-terminal that is not stable at low pH. Such a helix would be amphipathic.

While the rat brain sodium channel protein II has not been detected in rat skeletal muscles (Gordon et al., 1987), the region of homology with myotoxin *a is* especially unique with the conserved cysteines so critical to myotoxin *a*'s structure. Though this region is only 10 residues long, it encompasses 24% of myotoxin *a*'s sequence. Future homology searches in this area seem well warranted.

While it is purely speculative to define 2 domains in myotoxin *a*, bipolar regions of homology, Ca^{++} inhibition, and perhaps structure suggest this as a viable possibility. In such a model, one might envision myotoxin *a* with a N-terminal α-helix that anchors to the membrane surface while a region near the C-terminus with charged side chains pointing away from aromatic rings along a portion of anti-parallel β-sheet interacts with a sodium channel protein to effect an uncontrolled influx of Na^+ into the muscle cell.

Much additional investigation is needed to determine how myotoxin *a* functions and how that function relates to its structure. A highly defined tertiary structure, however, should be solvable within the near future through improved purification and ^{1}H-NMR techniques.

REFERENCES

Aird, S. D., Kruggel, W. G., & Kaiser, I. I. (1991) *Toxicon 29*, 265-268.

Allen, H. R., Tucker, R. K., & Geren C. R. (1986) *Toxicon 24*, 553-558.

Bailey, G. S., Lee, J., & Tu, A. T. (1979) *J. Biol. Chem. 254*, 8922-8926.

Baker, B., Utaisincharoen, P., & Tu, A. T. (1992) *Arch. Biochem. Biophys. 298*, 325-331.

Bax, A., & Davis, D. G. (1985) *J. Magn. Reson. 65*, 355-360.

Bieber, A. L., McParland, R. H., & Becker, R. R. (1987) *Toxicon 25*, 677-680.

Billeter, M., Braun, W., & Wüthrich, K. (1982) *J. Mol. Biol. 155*, 321-346.

Cameron, D. L., & Tu, A. T. (1977) *Biochemistry 16*, 2546-2553.

Cameron, D. L., & Tu, A. T. (1978) *Biochim. Biophys. Acta 532*, 147-154.

Chou, P. Y., & Fasman, G. D. (1974) *Biochemistry 13*, 222-245.

Conti, M. A. B., & Laure, C. J. (1988) in *Anais do XII Simpósio Anual da ACIESP*, Vol. 1, p 173, Campinas, São Paulo.

Creighton, T. E. (1993) *Proteins*, p 226, Freeman, New York.

Crippen, G. M. (1977) *J. Comput. Phys. 24*, 96-107.

Englander, S. W., & Wand, A. J. (1987) *Biochemistry 26*, 5953-5858.

Engle, C. M., Becker, R. R., Bailey, T., & Bieber, A. L. (1983) *J. Toxicol.-Toxin Rev. 2*, 267-283.

Fox, J. W., Elzinga, M., & Tu, A. T. (1979) *Biochemistry 18*, 678-684.

Garnier, J., Osguthorpe, D. J., & Robson, B. (1978) *J. Mol. Biol. 120*, 97-120.

Gleason, M., Odell, G., & Ownby, C. L. (1983) *J. Toxic.-Toxin Rev. 2*, 235-265.

Glenn, J. L., & Straight, R. C. (1982) in *Rattlesnake Venoms: Their Actions and Treatment* (Tu, A. T., Ed.) pp 66-111, Marcel Dekker, Inc., New York.

Gordon, D., Merrick, D., Auld, V., Dunn, R., Goldin, A. L., Davidson, N., & Catterall, W. A. (1987) *Proc. Natl. Acad. Sci. U.S.A. 84*, 8682-8686.

Griffin, P. R., & Aird, S. D. (1990) *FEBS Lett. 274*, 43-47.

Güntert, P., Braun, W., & Wüthrich, K. (1991a) *J. Mol. Biol. 217*, 517-530.

Güntert, P., Qian, Y. Q., Otting, G., Müller, M., Gehring, W., & Wüthrich, K. (1991b) *J. Mol. Biol. 217*, 531-540.

Harrison, R. K., & Stein, R. L. (1990) *Biochemistry 29*, 1684-1689.

Havel, T. F., & Wüthrich, K. (1985) *J. Mol. Biol. 182*, 281-294.

Hayes, C. E. (1984) Ph.D. Dissertation, Arizona State University, Tempe, AZ.

Henderson, J. T. (1986) Ph.D. Dissertation, Arizona State University, Tempe, AZ.

Henderson, J. T., & Bieber, A. L. (1987) in *Progress in Venom Research* (Gopalakrishnakone, P., & Tan, K., Eds.) pp 146-156, Republic of Signapore.

Henderson, J. T., Nieman, R. A., & Bieber, A. L. (1987) *Biochim. Biophys. Acta 914*, 152-161.

Hong, S. J., & Chang, C. C. (1985) *Toxicon 23*, 927-937.

Hsu, V. L., Handschumacher, R. E., & Armitage, I. M. (1990) *J. Am. Chem. Soc. 112*, 6745-6747.

Klauber, L. M. (1982) *Rattlesnakes: Their Habits, Life Histories, and Influence on Mankind*, Chapters 5, 6 and 10, University of California Press, Berkeley.

Kumar, A., Ernst, R. R., & Wüthrich, K. (1980) *Biochem. Biophys. Res. Comm. 95*, 1-6.

Laure, C. J. (1975) *Hoppe-Seyler's Z. Physiol. Chem. 356*, 213-215.

Maeda, N., Tamiya, N., Pattabhiraman, T. R., & Russell, F. E. (1978) *Toxicon 16*, 431-441.

Meadows, R. P., Kaluarachchi, K., Post, C. B., & Gorenstein, D. G. (1991) *Bull. Magn. Reson. 13*, 22-48.

Mebs, D., & Ownby, C. L. (1990) *Pharmac. Ther. 48*, 223-236.

Ménez, A. (1991) in *Snake Toxins* (Harvey, A. L., Ed.) pp 70-71, Pergamon Press, New York.

Metzler, W. J., Hare, D. R., & Pardi, A. (1989) *Biochemistry 28*, 7045-7052.

Misra, R. (1991) B.S. Honors Thesis, Arizona State University, Tempe, AZ.

Murchison, H. A. (1989) Ph.D. Dissertation, Arizona State University, Tempe, AZ.

Nerdal, W., Hare, D. R., & Reid, B. R. (1989) *Biochemistry 28*, 10008-10021.

Ownby, C. L. (1982) in *Rattlesnake Venoms: Their Actions and Treatment* (Tu, A. T., Ed.) pp 163-191, Marcel Dekker, Inc., New York.

Ownby, C. L., Aird, S. D., & Kaiser, I. I. (1988) *Toxicon 26*, 319-323.

Rance, M., Sorenson, O. W., Bodenhausen, G., Wagner, G., Ernst, R. R., & Wüthrich, K. (1983) *Biochem. Biophys. Res. Comm. 117*, 479-485.

Rawn, J. D. (1989) *Biochemistry*, p 77, Prentice Hall, Englewood Cliffs, New Jersey.

Samejima, Y., Aoki, Y., & Mebs, D. (1991) *Toxicon 29*, 461-468.

Smith, L. A., & Schmidt, J. J. (1990) *Toxicon 28*, 575-585.

States, D. J., Haberkorn, R. A., & Ruben, D. J. (1982) *J. Magn. Reson. 48*, 286-292.

Stocker, K. F. (1990) *Medical Use of Snake Venom Proteins*, pp 71-72, CRC Press, Boca Raton.

Teno, A. M., Vieira, C. A., Santoro, M. M., de Almeida Neves, A. G., & Giglio, J. R. (1990) *J. Biochem 107*, 821-825.

Tu, A. T. (1991) in *Reptile Venoms and Toxins* (Tu, A. T., Ed.) pp 317-329, Marcel Dekker, Inc., New York.

Tu, A. T., & Morita, M. (1983) *Br. J. exp. Path. 64*, 633-637.

Utaisincharoen, P., Baker, B., & Tu, A. T. (1991) *Biochemistry 30*, 8211-8216.

Wagner, G., Braun, W., Havel, T. F., Schaumann, T., Go, N., & Wüthrich, K. (1987) *J. Mol. Biol. 196*, 611-639.

Weber, P. L., Morrison, R. & Hare, D. (1988) *J. Mol. Biol. 204*, 483-487.

Wemmer, D. E. (1990) *Bull. Magn. Reson. 12*, 185-208.

Williamson, M. P., Havel, T. F., & Wüthrich, K. (1985) *J. Mol. Biol. 182*, 295-315.

Wüthrich, K. (1986) *NMR of Proteins and Nucleic Acids*, Parts I, II, and IV, Wiley-Interscience, New York.

Wüthrich, K., Billeter, M., & Braun, W. (1983) *J. Mol. Biol. 169*, 949-961.

APPENDIX A

DSPACE MACROS

zipref.mac

```
c**zipref.mac
;
; this macro will sequentially zip through your structure
; to fix the chirality, minimize and anneal in 4-D (if
; necessary) an increasing number of residues until the
; entire structure is treated as one
;
get/sym nstr "Number of structures to create:  "
get/sym refseq "Sequence to use:  "
get/sym refbmx "Bounds matrix to use:  "
get/sym fname "Filename for written structures:  "
;
set/log on %fname%
;
rea/seq %refseq%
rea/bmx %refbmx%
;
; set pertinent variables different from defaults
;
set n_cy 16  ;reasonable
set n_dim 4  ;enables 4th dimension
set 4d 1     ;enables 4D refinement
;
for refrun 1 %nstr%   ;main refinement run loop
  embed               ;embed in 4D space
  wri %fname%tmp      ;save orig embed structure(4D)
  fix/mirror          ;refine mirror image of embed first
  .zipfix   ;fixes indiv res ca chiraity w/ fix/mirror
  .newc_prochiral  ;swaps around most bad chiral centers
  fix/c            ;fixes any other bad chiral centers
  wri %fname%%refrun%apr  ;save pre-refinement mirror structure(4D)
  call zipit
  wri %fname%%refrun%a    ;save refined mirror structure(3D)
  cle/x
  rea %fname%tmp          ;get orig embed structure(3D)
  .zipfix
  .newc_prochiral
  fix/c
  wri %fname%%refrun%bpr  ;save pre-refinement structure(4D)
  call zipit
  wri %fname%%refrun%b    ;save refined structure(3D)
  cle/x
  rea/seq %refseq%        ;need sequence for next embed
endfor
end
;
zipit:
  for zipbig 1 %n_res%
    .cgrna %zipbig% (.2+.05*(%zipbig%-1)) ;sequential refine
  endfor
  set n_dim 3     ;goal is good 3D structure
  for rezip 1 10  ;10 final refinements or chiral fixes
    min           ;final refinements if chirality good
;   fix/c > cprob ;echoes any chirality problems fixed
;   if %cprob% .sne.     then ;if chirality probs were found
;     set n_dim 4            ;reset to 4D
;     set 4d 1
;     .cgrna 1 .3          ;zip refine individual residues
```

```
;      .cgrna %nres% (.3+.05*(%n_res%-1)) ;refine whole mol
;      set n_dim 3                        ;return to 3D
;    endif
  endfor
  return
```

zipfix.mac

```
c**zipfix.mac
; this macro looks at each residue, in order,
; calculates its signed volume for the ca chiral
; center, and if it is a negative value (D) changes
; it to positive (L) by using fix/mirror on only
; that residue
;
type "Performing ZipFix...
type " "
for hres 1 %n_res
  view/residue [%hres
  / f$residue{1} > cares
  if %cares% .sne. [gly then
;    .signvol ha n cb c
    set/sym signvol 0
    set/sym lha f$find{ha}
    set/sym ln f$find{n}
    set/sym lcb f$find{cb}
    set/sym lc f$find{c}
    / v$get{%lha%,41}
    / v$get{%ln%,42}
    / v$get{%lcb%,43}
    / v$get{%lc%,44}
    / v$sub{44,41,45} v$sub{44,42,46} v$sub{44,43,47}
    / v$cross{46,47,48} v$dot{45,48} > signvol
    ty " "
    ty "signed volume  of %cares% %hres% is " %signvol
    ty " "
    if %signvol% .lt. 0 then
      fix/mirror
      type "Fix/mirror performed on "%cares% %hres%
      type " "
    endif
  endif
endfor
view/all
dra
end
```

cgrna.mac

```
c**cgrna.mac
get/sym nres "Number of sequential residues per min:  "
get/sym pval "Acceptable total penalty value:  "
for res 1 (%n_res%-%nres%+1)
;
  if %nres% .eq. 1 then
    set/sym tres %nres%
    call ann
  endif
;
begin:
  set/sym ptry 0
```

```
  call min
  for %ptry% 1 20
  if %p_tot% .gt. %pval% then
    call ann
    call min
  endif
  endfor
  if %p_tot% .gt. %pval% then
    call bombout
  endif
  goto drwnfix
;
min:
  / (%res%+%nres%-1) > tres
  type " "
  type "Mimimizing residues %res% to %tres%"
  fix/chiral
  if %nres% .lt. 23 then
    for bitmin 1 (%nres%+3)
      min *[%res%:%tres%] 8.0 !{*[%res%:%tres%]}
    endfor
  else
    for settmin 1 25
      min *[%res%:%tres%] 8.0 !{*[%res%:%tres%]}
    endfor
  endif
  return
;
ann:
  type " "
  type "Penalty value exceeded -- annealing..."
  set wt_4d 0.1
  set shake 0
  for 4dhot 1 5
    anneal *[%res%:%tres%]
  endfor
  for 4dcool 1 5
    set wt_4d (%wt_4d%+.18)
    anneal *[%res%:%tres%]
  endfor
  return
;
bombout:
  type " "
  type "Could not achieve acceptable penalty value for %res%:%tres%."
  type %p_tot%
  type " "
  return
;
drwnfix:
  dr
  fix/chiral
;
endfor
end
```

wholeanne.mac

```
c**wholeanne.mac
set/sym trhs %n_res%
```

```
call anne
call mini
end
;
anne:
  type " "
  type "Performing whole molecule annealing..."
  set wt_4d 0.1
  set shake 0
  for 4dwarm 1 20
    anneal *[1:%trhs%]
  endfor
  for ghostmin 1 15
    min
  endfor
  for 4dchau 1 10
    anneal *[1:%trhs%]
  endfor
  for 4dcold 1 5
    set wt_4d (%wt_4d%+.18)
    anneal *[1:%trhs%]
  endfor
  return
;
mini:
  type " "
  type "...and whole molecule minimization."
  fix/chiral
  for pullmein 1 10
    min
  endfor
  return
irisgt 47% cat min50.mac
c**more50.mac
;performs min 50 times
for tightenup 1 50
  min
endfor
end
```

APPENDIX B

CONVERSION FILTERS

konconvert convert DIANA .cor coordinate output files to Quanta konnert input format

```
cat $1 | sed -e '/Q/ d
/DIANA/ d
/atoms/ d
s/xxx/yyy/g' > ee1
awk '{printf(" %2d   %4s  %-4s %8s %8s %8s%s",$3,$4,$2,$5,$6,$7,ORS)}'
<ee1 >ee2
cat ee2 | sed -e 's/ARG  HH1/ARG  HH11/g
/ASP  HD2/ d
s/ARG+/ARG /g
s/LYS+/LYS /g
s/ASP-/ASP /g
s/HIS+/HIS /g
s/CYSS/CYS /g
s/HB2/HB1/g
s/HB3/HB2/g
s/HD2/HD1/g
s/HD3/HD2/g
s/HG2/HG1/g
s/HG3/HG2/g
s/HE2/HE1/g
s/HE3/HE2/g
s/HN/H /g
s/xxx/yyy/g' > ee3
 awk '{printf("  %3s    %2d%-4s %8s0 %8s0 %8s0
0.00000%s",$2,$1,$3,$4,$5,$6,ORS)}' <ee3 >$2
```

assign.filter (edited) example of file that filters Felix 2.05 written entities replacing arbitrary numerical assignments with sequence specific assignments

```
cat $1 | sed -e 's/ca63/caK2/g
s/cb63/cbK2/g
s/nh63/nhK2/g
s/cdPIII/cdP13/g
s/cgPIII/cgP13/g
s/ca45/caK14/g
s/cb45/cbK14/g
s/nh64/nhK16/g
s/c?64/c?K16/g
s/ca41/caI17/g
s/c?36/c?I19/g
s/caPII/caP21/g
s/caPI/caP20/g
s/nh28/nhC30/g
s/nh29/nhK39/g
s/c?32/c?S41/g
s/2hWII/2hW32/g
s/7hWb/7hW34/g
s/2hWI/2hW34/g
s/neWI/neW34/g
s/cbWI/cbW34/g' > taw
awk '{printf("   %3s  %8s    %6s     %1s        % 6s  %8s    %6s     %s
%6s %5s%s",$1,$2,$3,$4,$5,$6,$7,$8,$9,$10,ORS)}' <taw >$2
rm taw
```

APPENDIX C

DIANA FILES

myoamod10.upl last version of original constraints

```
 4 CYSS
        CB      36 CYSS SG      3.10
        SG      36 CYSS CB      3.10
        SG      36 CYSS SG      2.10
 8 GLY
        HN       9 GLY  HN      3.50
 9 GLY
        HN      10 HIS+ HN      5.00
        HN      38 LYS+ HA      5.00
        QA      38 LYS+ HA      4.38
10 HIS+
        HN      10 HIS+ QB      3.76
        HN      11 CYSS HN      5.00
        HN      12 PHE  HE2     5.00
        HN      36 CYSS HA      5.00
        HN      37 CYSS HN      4.00
        HN      38 LYS+ HA      5.00
        HN      39 LYS+ HA      5.00
        HA      11 CYSS HN      3.50
        QB      11 CYSS HN      3.68
        QB      12 PHE  HE2     4.88
11 CYSS
        HN      11 CYSS QB      3.76
        HN      12 PHE  HN      5.00
        HA      12 PHE  HN      2.80
        HA      37 CYSS HN      5.00
        CB      30 CYSS SG      3.10
        QB      12 PHE  HN      4.38
        QB      13 PRO  QD      6.75
        SG      30 CYSS CB      3.10
        SG      30 CYSS SG      2.10
12 PHE
        HN      34 TRP  HA      6.00
        HN      35 LYS+ HN      3.50
        HN      36 CYSS HA      5.00
        HN      37 CYSS HN      5.00
        HE2     37 CYSS HN      5.00
13 PRO
        HA      34 TRP  HA      5.00
15 GLU
        HN      16 LYS+ HN      5.00
        HA      16 LYS+ HN      2.80
16 LYS+
        HN      17 ILE  HN      5.00
        HA      17 ILE  HN      2.80
        QB      17 ILE  HN      5.88
17 ILE
        HN      18 CYSS HN      5.00
        HA      18 CYSS HN      2.80
18 CYSS
        HN      19 ILE  HN      5.00
        HA      19 ILE  HN      2.80
        CB      37 CYSS SG      3.10
        QB      19 ILE  HN      4.38
        QB      24 ASP  QB      5.75
        SG      37 CYSS CB      3.10
        SG      37 CYSS SG      2.10
20 PRO
```

	HA	21	PRO	QD	3.32
21 PRO					
	HA	24	ASP	QB	5.88
22 SER					
	HN	22	SER	QB	3.77
	HN	23	SER	HN	5.00
	HA	23	SER	HN	3.50
	QB	23	SER	HN	3.68
23 SER					
	HN	24	ASP	HN	3.50
	HA	38	LYS+	HN	5.00
	QB	38	LYS+	HN	5.88
24 ASP					
	HN	25	LEU	HN	5.00
	HA	25	LEU	HN	2.80
	HA	26	GLY	HN	5.00
	HA	37	CYSS	HA	4.00
	HA	38	LYS+	HN	5.00
	QB	25	LEU	HN	3.68
	QB	37	CYSS	QB	4.75
25 LEU					
	HN	26	GLY	HN	3.50
	HN	26	GLY	QA	5.88
	HN	27	LYS+	HN	5.00
	HN	37	CYSS	HA	4.00
	HA	26	GLY	HN	3.50
	QB	26	GLY	HN	5.88
26 GLY					
	HN	27	LYS+	HN	6.00
	QA	27	LYS+	HN	3.13
27 LYS+					
	HN	27	LYS+	QB	3.77
	HN	28	MET	HN	5.00
	HA	28	MET	HN	3.50
	HA	36	CYSS	HN	5.00
	QB	28	MET	HN	3.68
30 CYSS					
	HN	30	CYSS	QB	3.76
	HN	31	ARG+	HN	5.00
	HA	31	ARG+	HN	2.80
	QB	31	ARG+	HN	4.38
31 ARG+					
	HA	32	TRP	HN	3.50
32 TRP					
	HA	34	TRP	HN	5.00
33 LYS+					
	HA	34	TRP	HN	3.50
	QB	34	TRP	HN	5.88
34 TRP					
	HN	34	TRP	QB	3.76
	HN	35	LYS+	HN	5.00
	HA	35	LYS+	HN	3.50
35 LYS+					
	HN	36	CYSS	HN	6.00
36 CYSS					
	HN	36	CYSS	QB	3.76
	HN	37	CYSS	HN	5.00
	HA	37	CYSS	HN	2.80
37 CYSS					

```
          HN      37 CYSS QB       3.76
          HN      38 LYS+ HN       5.00
          HA      38 LYS+ HN       2.80
          QB      38 LYS+ HN       4.38
 40 GLY
          HN      41 SER  HN       5.00
          HN      41 SER  HA       6.00
 41 SER
          HA      42 GLY  HN       3.50

myoamod12.upl DIANA modified version of myoamod11.upl
  4 CYSS
          CB      36 CYSS SG       3.10
          SG      36 CYSS CB       3.10
          SG      36 CYSS SG       2.10
  7 LYS+
          HN       7 LYS+ QB       3.54
  8 GLY
          HN       9 GLY  HN       3.00
  9 GLY
          HA1     38 LYS+ HA       4.00
          HA2     38 LYS+ HA       4.00
          QA      38 LYS+ HA       3.43
 10 HIS+
          HN      12 PHE  HE2      4.00
          HN      37 CYSS HN       3.00
          QB      11 CYSS HN       3.68
          QB      12 PHE  HE2      3.88
 11 CYSS
          HA      12 PHE  HN       3.00
          HA      36 CYSS HA       3.00
          HA      37 CYSS HN       4.00
          CB      30 CYSS SG       3.10
          QB      12 PHE  HN       5.88
          SG      30 CYSS CB       3.10
          SG      30 CYSS SG       2.10
 12 PHE
          HN      12 PHE  QB       3.53
          HN      35 LYS+ HN       3.00
          HN      36 CYSS HA       5.00
          HE2     37 CYSS HN       4.00
          HE2     37 CYSS HB2      4.00
          HE2     37 CYSS HB3      4.00
          HE2     37 CYSS QB       3.43
 13 PRO
          HA      34 TRP  HA       4.00
 14 LYS+
          HN      15 GLU  HN       3.00
          HN      34 TRP  HA       3.00
          HN      35 LYS+ HN       4.00
          HA      16 LYS+ HN       4.00
 15 GLU
          HN      15 GLU  QG       4.88
          HN      16 LYS+ HN       3.00
 16 LYS+
          HN      16 LYS+ QB       3.54
          HA      18 CYSS HN       3.00
 17 ILE
          HN      17 ILE  HB       3.00
```

```
18 CYSS
        CB      37 CYSS SG      3.10
        QB      24 ASP  QB      4.75
        SG      37 CYSS CB      3.10
        SG      37 CYSS SG      2.10
19 ILE
        HN      19 ILE  HB      3.00
        HA      20 PRO  HA      3.00
21 PRO
        HA      24 ASP  QB      4.88
22 SER
        HN      22 SER  HB2     4.00
        HN      22 SER  HB3     4.00
        HN      22 SER  QB      3.43
        HN      23 SER  HN      3.00
23 SER
        HN      23 SER  HB2     4.00
        HN      23 SER  HB3     4.00
        HN      23 SER  QB      3.43
        HN      24 ASP  HN      4.00
        QB      38 LYS+ HN      4.88
24 ASP
        HN      24 ASP  QB      3.54
        HA      26 GLY  HN      4.00
        HA      37 CYSS HA      3.00
        QB      37 CYSS QB      4.75
25 LEU
        HN      25 LEU  QB      3.54
        HN      26 GLY  HN      3.00
        HN      37 CYSS HA      3.00
26 GLY
        QA      27 LYS+ HN      3.22
27 LYS+
        HN      28 MET  HN      4.00
        HA      27 LYS+ QG      3.58
        HA      36 CYSS HN      4.00
30 CYSS
        HN      30 CYSS HB2     4.00
        HN      30 CYSS HB3     4.00
        HN      30 CYSS QB      3.43
        HA      31 ARG+ HN      3.00
31 ARG+
        HN      31 ARG+ QB      3.55
        HN      34 TRP  QB      3.88
        HB2     34 TRP  HD1     4.00
        HB3     34 TRP  HD1     4.00
        QB      31 ARG+ HE      4.88
        QB      32 TRP  HN      3.88
        QB      34 TRP  HD1     3.43
32 TRP
        HN      32 TRP  HE3     5.00
        HN      33 LYS+ HN      4.00
        HA      33 LYS+ HN      3.00
        QB      32 TRP  HD1     3.38
        HE3     33 LYS+ HN      4.00
33 LYS+
        HN      34 TRP  HN      3.00
        QB      34 TRP  HD1     4.88
34 TRP
```

```
          HN      34 TRP  HD1      3.00
          HA      35 LYS+ HN       3.00
          HB2     35 LYS+ HN       4.00
          HB3     35 LYS+ HN       4.00
          QB      34 TRP  HD1      3.38
          QB      35 LYS+ HN       3.43
 37 CYSS
          HN      37 CYSS HB2      4.00
          HN      37 CYSS HB3      4.00
          HN      37 CYSS QB       3.43
          QB      38 LYS+ HN       4.88
 38 LYS+
          HN      38 LYS+ HB2      4.00
          HN      38 LYS+ HB3      4.00
          HN      38 LYS+ QB       3.43
 39 LYS+
          HN      39 LYS+ QB       3.54
```

myoa8.ovw (edited) structures built with myoamod10.upl (last version of original constraints)
Overview:

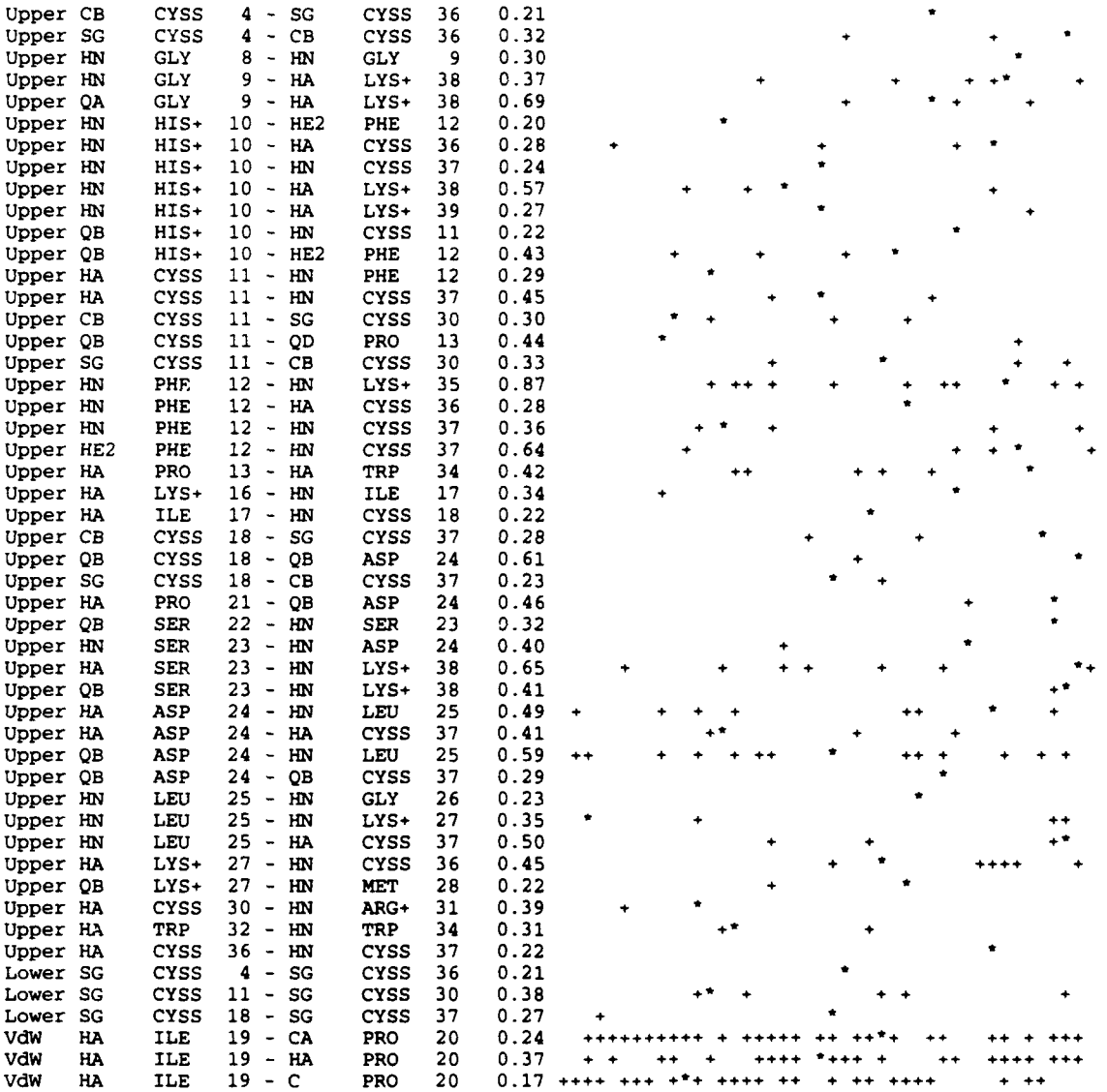
```
Number of accepted structures  :        536 (536 structures started)
Residue range for upper limits :         42
                lower limits :         42
                van der Waals:         42
Cutoff for upper limits        :       0.20 A
           lower limits        :       0.20 A
           van der Waals       :       0.10 A
           angle constraints   :       5.00 deg
CPU time                       :     568.83 min
CPU time per structure         :       1.06 min
Average number of iterations   :       2460
```

struct	target	upper limits			lower limits			van der Waals			torsion angles		
	function	#	sum	max	#	sum	max	#	sum	max	#	sum	max
1 306	0.71	0	0.8	0.19	0	0.0	0.03	7	2.9	0.34	0	0.0	0.0
2 173	0.89	2	1.9	0.40	0	0.0	0.02	7	2.8	0.21	0	0.0	0.0
3 159	0.94	2	1.7	0.35	0	0.1	0.07	8	3.3	0.31	0	0.0	0.0
4 490	0.94	0	1.9	0.20	1	0.3	0.24	7	3.8	0.23	0	0.0	0.0
5 38	1.04	1	1.6	0.21	0	0.3	0.15	8	3.2	0.41	0	0.0	0.0
6 90	1.06	2	1.9	0.33	0	0.2	0.10	7	4.2	0.33	0	0.0	0.0
7 437	1.08	0	1.1	0.18	0	0.2	0.12	13	4.6	0.28	0	0.0	0.0
8 149	1.16	0	1.5	0.19	0	0.1	0.06	14	4.9	0.24	0	0.0	0.0
9 343	1.31	4	2.5	0.44	0	0.1	0.05	10	3.5	0.36	0	0.0	0.0
10 513	1.35	2	2.2	0.30	0	0.4	0.18	13	4.7	0.27	0	0.0	0.0
11 15	1.50	2	2.4	0.58	0	0.2	0.19	14	4.6	0.22	0	0.0	0.0
12 198	1.52	5	2.9	0.44	1	0.5	0.29	8	3.5	0.31	0	0.0	0.0
13 258	1.53	4	2.1	0.35	1	0.6	0.38	13	4.6	0.26	0	0.0	0.0
14 154	1.58	5	2.6	0.41	0	0.1	0.05	14	4.6	0.27	0	0.0	0.0
15 327	1.71	5	2.7	0.38	0	0.1	0.04	14	4.7	0.31	0	0.0	0.0
16 294	1.71	3	2.3	0.35	1	0.4	0.30	14	5.2	0.34	0	0.0	0.0
17 246	1.72	3	1.7	0.37	0	0.0	0.01	17	5.6	0.29	0	0.0	0.0
18 164	1.83	7	3.3	0.51	0	0.0	0.02	11	4.0	0.35	0	0.0	0.0
19 118	1.96	3	3.2	0.57	0	0.3	0.09	18	5.0	0.25	0	0.0	0.0
20 512	1.98	0	1.1	0.19	0	0.1	0.10	23	6.2	0.28	0	0.0	0.0
21 450	2.06	2	1.9	0.30	0	0.0	0.03	9	5.2	0.57	0	0.0	0.0
22 137	2.16	4	2.4	0.45	0	0.1	0.06	19	6.1	0.37	0	0.0	0.0
23 214	2.17	5	3.6	0.59	1	0.5	0.27	15	4.4	0.39	0	0.0	0.0
24 77	2.20	3	2.3	0.32	1	0.4	0.21	16	7.1	0.41	0	0.0	0.0
530 353	67.07	17	13.9	1.49	0	0.1	0.13	191	69.6	1.49	0	0.0	0.0
531 141	70.19	18	11.9	1.28	0	0.2	0.15	178	74.8	2.10	0	0.0	0.0
532 469	80.13	25	18.3	2.57	0	0.0	0.00	237	81.7	1.17	0	0.0	0.0
533 516	80.76	19	15.4	1.74	3	1.7	0.72	239	87.8	1.53	0	0.0	0.0

534	318	81.61	23	15.8	1.30	0	0.0	0.00	234	89.3	2.21	0	0.0	0.0
535	165	84.55	21	14.7	1.49	0	0.2	0.15	191	82.0	2.11	0	0.0	0.0
536	448	92.05	22	19.2	1.73	0	0.0	0.00	230	89.9	1.58	0	0.0	0.0

Constraint violation and hydrogen bond overview (structures ordered):

Cutoff for target function : 3.00E+00
Number of structures included : 44
Number of violated constraints : 468
Number of consistent violations: 0
Maximal hydrogen bond length : 2.40 A
Maximal hydrogen bond angle : 35.00 deg
Number of hydrogen bonds : 146
Number of consistent H-bonds : 0



Upper	CB	CYSS	4	SG	CYSS	36	0.21	
Upper	SG	CYSS	4	CB	CYSS	36	0.32	
Upper	HN	GLY	8	HN	GLY	9	0.30	
Upper	HN	GLY	9	HA	LYS+	38	0.37	
Upper	QA	GLY	9	HA	LYS+	38	0.69	
Upper	HN	HIS+	10	HE2	PHE	12	0.20	
Upper	HN	HIS+	10	HA	CYSS	36	0.28	
Upper	HN	HIS+	10	HN	CYSS	37	0.24	
Upper	HN	HIS+	10	HA	LYS+	38	0.57	
Upper	HN	HIS+	10	HA	LYS+	39	0.27	
Upper	QB	HIS+	10	HN	CYSS	11	0.22	
Upper	QB	HIS+	10	HE2	PHE	12	0.43	
Upper	HA	CYSS	11	HN	PHE	12	0.29	
Upper	HA	CYSS	11	HN	CYSS	37	0.45	
Upper	CB	CYSS	11	SG	CYSS	30	0.30	
Upper	QB	CYSS	11	QD	PRO	13	0.44	
Upper	SG	CYSS	11	CB	CYSS	30	0.33	
Upper	HN	PHE	12	HN	LYS+	35	0.87	
Upper	HN	PHE	12	HA	CYSS	36	0.28	
Upper	HN	PHE	12	HN	CYSS	37	0.36	
Upper	HE2	PHE	12	HN	CYSS	37	0.64	
Upper	HA	PRO	13	HA	TRP	34	0.42	
Upper	HA	LYS+	16	HN	ILE	17	0.34	
Upper	HA	ILE	17	HN	CYSS	18	0.22	
Upper	CB	CYSS	18	SG	CYSS	37	0.28	
Upper	QB	CYSS	18	QB	ASP	24	0.61	
Upper	SG	CYSS	18	CB	CYSS	37	0.23	
Upper	HA	PRO	21	QB	ASP	24	0.46	
Upper	QB	SER	22	HN	SER	23	0.32	
Upper	HN	SER	23	HN	ASP	24	0.40	
Upper	HA	SER	23	HN	LYS+	38	0.65	
Upper	QB	SER	23	HN	LYS+	38	0.41	
Upper	HA	ASP	24	HN	LEU	25	0.49	
Upper	HA	ASP	24	HA	CYSS	37	0.41	
Upper	QB	ASP	24	HN	LEU	25	0.59	
Upper	QB	ASP	24	QB	CYSS	37	0.29	
Upper	HN	LEU	25	HN	GLY	26	0.23	
Upper	HN	LEU	25	HN	LYS+	27	0.35	
Upper	HN	LEU	25	HA	CYSS	37	0.50	
Upper	HA	LYS+	27	HN	CYSS	36	0.45	
Upper	QB	LYS+	27	HN	MET	28	0.22	
Upper	HA	CYSS	30	HN	ARG+	31	0.39	
Upper	HA	TRP	32	HN	TRP	34	0.31	
Upper	HA	CYSS	36	HN	CYSS	37	0.22	
Lower	SG	CYSS	4	SG	CYSS	36	0.21	
Lower	SG	CYSS	11	SG	CYSS	30	0.38	
Lower	SG	CYSS	18	SG	CYSS	37	0.27	
VdW	HA	ILE	19	CA	PRO	20	0.24	
VdW	HA	ILE	19	HA	PRO	20	0.37	
VdW	HA	ILE	19	C	PRO	20	0.17	

Pairwise RMSDs (structures ordered):

Number of backbone atoms	:	102
Number of heavy atoms	:	273
Residues considered	:	6..39
Local RMSD segment length	:	3 residues
Mean global backbone RMSD	:	3.71 +/- 0.76 A (1.86..6.58 A)
Mean global heavy atom RMSD	:	5.58 +/- 0.76 A (3.74..8.52 A)

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
1	306	2.25	2.80	2.90	2.96	3.42	2.87	2.50	2.84	2.04	2.69	3.05	3.61	1.86	2.91
2	4.09	173	3.65	3.38	2.75	3.29	3.58	3.15	3.31	2.37	3.67	2.83	3.94	2.40	2.93
3	4.31	5.81	159	2.95	3.54	3.65	3.13	2.67	3.69	3.12	2.57	3.26	4.09	3.04	3.19
4	4.42	5.21	4.62	490	2.64	2.78	2.61	2.93	3.84	2.35	2.51	3.38	4.12	3.07	2.87
5	4.87	4.85	6.00	4.58	38	2.85	2.70	3.28	3.06	1.87	3.04	3.43	4.25	2.53	2.80
6	5.05	5.33	5.39	4.53	4.92	90	3.07	3.15	4.17	2.87	3.35	2.91	3.84	3.30	2.82
7	4.44	5.73	4.34	4.23	4.82	4.86	437	3.23	3.09	2.50	2.64	3.63	3.64	3.23	2.87
8	4.80	5.32	4.76	4.64	5.60	5.24	5.00	149	3.79	2.74	2.78	2.71	4.04	2.66	3.11
9	4.59	4.58	6.09	5.80	4.75	6.18	4.95	5.92	343	3.21	3.41	3.85	4.32	3.19	3.44
10	3.96	4.19	5.57	4.52	4.16	5.07	4.63	5.08	4.77	513	2.84	3.14	3.90	1.91	2.56
11	4.46	5.63	5.00	4.66	4.68	4.86	4.45	4.92	5.09	4.78	15	3.49	3.95	3.17	3.37
12	4.77	4.83	5.32	5.24	5.77	4.53	5.61	4.74	5.71	5.02	5.52	198	4.52	3.26	2.79
13	5.16	5.83	5.53	5.73	6.40	6.01	5.36	5.55	6.32	5.83	6.05	6.18	258	3.92	4.06
14	3.77	3.98	5.13	4.56	4.43	5.19	5.26	4.90	5.07	3.96	5.17	4.71	5.54	154	2.99
15	4.74	5.02	5.39	4.90	4.68	4.57	5.12	5.48	5.48	5.00	5.11	5.06	5.75	4.83	327

Residue	local RMSD/segment		global RMSD/residue	
	backbone	heavy atoms	backbone	heavy atoms
1	0.00	0.00	13.06	14.46
2	0.89	2.79	10.87	12.41
3	0.91	2.29	8.41	9.67
4	0.86	2.40	5.98	5.71
5	0.90	2.35	6.21	8.24
6	0.85	2.69	5.86	7.93
7	0.82	2.10	4.86	6.87
8	0.83	1.78	3.80	3.96
9	0.78	1.61	2.48	2.59
10	0.77	1.68	2.28	4.80
11	0.64	1.90	2.12	2.51
12	0.60	1.48	1.99	3.75
13	0.73	2.25	2.73	3.36
14	0.85	2.25	3.46	5.71
15	0.81	2.71	3.68	5.76
16	0.85	2.40	3.61	5.98
17	0.82	2.18	3.31	4.85
18	0.75	1.92	2.42	2.62
19	0.66	1.42	3.52	4.45
20	0.32	1.03	3.71	4.25
21	0.63	1.19	3.21	3.84
22	0.74	1.45	3.54	4.29
23	0.67	1.57	2.90	3.52
24	0.82	1.92	2.43	3.61
25	0.74	1.79	3.35	5.02
26	0.84	2.12	3.60	3.78
27	0.87	2.20	3.29	5.79
28	0.91	2.51	4.08	5.84
29	0.86	2.06	4.22	5.63
30	0.83	2.33	3.52	3.69
31	0.86	2.59	3.72	7.24
32	0.80	3.06	3.67	6.72
33	0.79	3.14	3.98	6.26
34	0.81	2.57	3.04	6.14
35	0.69	2.22	2.46	4.cat

MYOA9.OVW (edited) structures from new constraints (myoamod12.upl)

Overview:

Number of accepted structures	:	999 (999 structures started)
Residue range for upper limits	:	42
lower limits	:	42
van der Waals	:	42
Cutoff for upper limits	:	0.20 A
lower limits	:	0.20 A
van der Waals	:	0.20 A
angle constraints	:	5.00 deg
CPU time	:	1034.44 min
CPU time per structure	:	1.04 min
Average number of iterations	:	2539

struct	target function	upper limits #	upper limits sum	upper limits max	lower limits #	lower limits sum	lower limits max	van der Waals #	van der Waals sum	van der Waals max	torsion angles #	torsion angles sum	torsion angles max
1 484	0.49	0	0.4	0.06	0	0.1	0.04	1	2.7	0.26	0	0.0	0.0
2 653	0.55	0	1.3	0.20	0	0.2	0.14	1	2.2	0.22	0	0.0	0.0
3 796	0.63	1	1.3	0.51	0	0.0	0.04	0	1.6	0.17	0	0.0	0.0
4 902	0.67	0	1.4	0.18	0	0.0	0.02	2	2.9	0.24	0	0.0	0.0
5 342	0.76	0	1.2	0.18	0	0.0	0.01	2	3.3	0.22	0	0.0	0.0
6 755	0.85	4	1.9	0.46	0	0.1	0.06	0	2.1	0.17	0	0.0	0.0
7 663	1.07	2	1.8	0.28	0	0.2	0.11	0	4.6	0.20	0	0.0	0.0
8 479	1.07	3	2.6	0.30	0	0.1	0.05	1	3.9	0.20	0	0.0	0.0
9 766	1.09	2	1.9	0.33	0	0.1	0.08	3	3.2	0.35	0	0.0	0.0
10 157	1.12	2	1.9	0.65	0	0.0	0.02	0	1.8	0.15	0	0.0	0.0
11 562	1.18	0	1.6	0.18	0	0.1	0.06	5	4.8	0.27	0	0.0	0.0
12 365	1.28	4	2.6	0.34	0	0.3	0.10	1	4.5	0.22	0	0.0	0.0
13 240	1.38	4	2.7	0.64	1	0.2	0.25	1	3.1	0.25	0	0.0	0.0
14 313	1.46	4	2.7	0.35	0	0.5	0.16	2	4.3	0.25	0	0.0	0.0
15 836	1.67	1	1.8	0.62	1	0.3	0.26	5	4.0	0.31	0	0.0	0.0
16 465	1.84	6	3.4	0.42	1	0.5	0.28	1	4.8	0.21	0	0.0	0.0
17 722	1.91	5	2.6	0.37	0	0.2	0.13	6	5.7	0.35	0	0.0	0.0
18 68	1.91	4	2.3	0.37	1	0.7	0.33	3	6.1	0.44	0	0.0	0.0
19 501	1.95	3	2.4	0.42	0	0.4	0.19	5	5.5	0.26	0	0.0	0.0
20 156	1.97	2	3.2	0.62	0	0.2	0.11	2	5.3	0.29	0	0.0	0.0
21 670	2.01	1	1.8	0.57	0	0.1	0.09	7	5.1	0.42	0	0.0	0.0
22 881	2.08	7	3.4	0.41	1	0.5	0.23	1	5.2	0.37	0	0.0	0.0
23 114	2.23	4	3.3	0.45	1	0.2	0.20	5	6.4	0.30	0	0.0	0.0
24 273	2.24	5	3.0	0.49	0	0.4	0.13	6	5.9	0.36	0	0.0	0.0
25 149	2.31	5	3.2	0.53	1	0.4	0.30	5	6.1	0.35	0	0.0	0.0
990 553	109.86	30	20.3	1.81	0	0.0	0.00	177	108.5	1.97	0	0.0	0.0
991 863	112.52	24	18.7	1.88	0	0.0	0.00	192	114.8	1.66	0	0.0	0.0
992 656	113.91	34	27.3	1.48	0	0.4	0.18	186	111.2	1.59	0	0.0	0.0
993 782	115.73	27	18.6	1.86	1	0.4	0.34	169	108.9	1.95	0	0.0	0.0
994 833	122.41	23	15.9	1.86	1	0.5	0.29	187	118.2	2.28	0	0.0	0.0
995 506	122.51	32	25.7	1.94	1	0.6	0.43	202	118.9	1.93	0	0.0	0.0
996 207	126.10	33	27.6	2.01	0	0.0	0.05	179	104.3	2.18	0	0.0	0.0
997 376	140.28	25	18.8	1.66	1	0.5	0.35	195	129.5	1.95	0	0.0	0.0
998 991	140.59	32	24.8	2.26	0	0.1	0.07	267	143.9	1.65	0	0.0	0.0
999 432	178.70	34	28.9	1.72	0	0.1	0.07	257	159.9	2.52	0	0.0	0.0
Average	32.75	19	12.3	1.29	0	0.4	0.21	58	35.4	1.09	0	0.0	0.0
+/-	23.74	7	4.8	0.45	1	0.3	0.14	38	21.8	0.47	0	0.0	0.0
Minimum	0.49	0	0.4	0.06	0	0.0	0.00	0	1.6	0.15	0	0.0	0.0
Maximum	178.70	35	28.9	3.02	4	2.1	1.08	267	159.9	2.77	0	0.0	0.0

Constraint violation and hydrogen bond overview (structures ordered):

Cutoff for target function	:	2.00E+00
Number of structures included	:	20

Number of violated constraints : 62
Number of consistent violations: 0
Maximal hydrogen bond length : 2.40 A
Maximal hydrogen bond angle : 35.00 deg
Number of hydrogen bonds : 66
Number of consistent H-bonds : 0

								max 1 5 10 15 20
Upper	HA2	GLY	9 -	HA	LYS+	38	0.37	
Upper	QA	GLY	9 -	HA	LYS+	38	0.30	
Upper	QB	HIS+	10 -	HE2	PHE	12	0.28	
Upper	CB	CYSS	11 -	SG	CYSS	30	0.41	
Upper	SG	CYSS	11 -	CB	CYSS	30	0.22	
Upper	HN	PHE	12 -	HN	LYS+	35	0.24	
Upper	HE2	PHE	12 -	HB2	CYSS	37	0.22	
Upper	HN	LYS+	14 -	HN	GLU	15	0.34	
Upper	HN	LYS+	14 -	HA	TRP	34	0.27	
Upper	HA	LYS+	14 -	HN	LYS+	16	0.26	
Upper	HA	LYS+	16 -	HN	CYSS	18	0.33	
Upper	HN	ILE	17 -	HB	ILE	17	0.20	
Upper	CB	CYSS	18 -	SG	CYSS	37	0.28	
Upper	SG	CYSS	18 -	CB	CYSS	37	0.27	
Upper	HN	ILE	19 -	HB	ILE	19	0.24	
Upper	HN	LEU	25 -	HA	CYSS	37	0.21	
Upper	HN	LYS+	27 -	HN	MET	28	0.35	
Upper	HA	LYS+	27 -	HN	CYSS	36	0.62	
Upper	HN	CYSS	30 -	QB	CYSS	30	0.28	
Upper	HA	CYSS	30 -	HN	ARG+	31	0.65	
Upper	HN	ARG+	31 -	QB	TRP	34	0.42	
Upper	QB	ARG+	31 -	HN	TRP	32	0.51	
Upper	HN	TRP	32 -	HN	LYS+	33	0.62	
Upper	HA	TRP	32 -	HN	LYS+	33	0.64	
Upper	QB	LYS+	33 -	HD1	TRP	34	0.26	
Upper	HN	TRP	34 -	HD1	TRP	34	0.22	
Lower	SG	CYSS	11 -	SG	CYSS	30	0.28	
Lower	SG	CYSS	18 -	SG	CYSS	37	0.33	
VdW	HA	HIS+	5 -	HD1	HIS+	5	0.22	
VdW	O	HIS+	5 -	HG3	LYS+	6	0.20	
VdW	HG3	LYS+	6 -	O	LYS+	6	0.20	
VdW	HA1	GLY	9 -	C	LYS+	38	0.44	
VdW	HN	HIS+	10 -	HE2	PHE	12	0.24	
VdW	HN	HIS+	10 -	C	CYSS	37	0.22	
VdW	O	HIS+	10 -	CB	CYSS	11	0.25	
VdW	O	HIS+	10 -	HB2	CYSS	11	0.23	
VdW	HB3	PHE	12 -	HD2	PRO	13	0.21	
VdW	HD1	PHE	12 -	HD11	ILE	17	0.22	
VdW	C	PRO	13 -	HN	GLU	15	0.21	
VdW	O	PRO	13 -	CB	LYS+	14	0.23	
VdW	HB3	LYS+	14 -	CD1	ILE	17	0.23	
VdW	C	LYS+	14 -	HN	LYS+	16	0.26	
VdW	HN	GLU	15 -	HN	LYS+	16	0.25	
VdW	HN	LYS+	16 -	HD12	ILE	17	0.35	
VdW	HA	ILE	19 -	CA	PRO	20	0.23	
VdW	HA	ILE	19 -	HA	PRO	20	0.35	
VdW	HG23	ILE	19 -	C	ILE	19	0.24	
VdW	O	ASP	24 -	CB	LEU	25	0.22	
VdW	O	LEU	25 -	C	GLY	26	0.27	
VdW	HB3	ASP	29 -	HN	CYSS	30	0.20	
VdW	O	CYSS	30 -	C	ARG+	31	0.22	
VdW	HB3	ARG+	31 -	HN	TRP	32	0.21	
VdW	C	ARG+	31 -	HN	LYS+	33	0.24	
VdW	HN	TRP	32 -	HN	LYS+	33	0.29	
VdW	CG	TRP	32 -	O	TRP	32	0.26	
VdW	C	TRP	32 -	HN	TRP	34	0.27	
VdW	HN	LYS+	33 -	HN	TRP	34	0.24	
VdW	HG3	LYS+	33 -	C	LYS+	33	0.22	
VdW	O	LYS+	33 -	O	TRP	34	0.26	
VdW	N	TRP	34 -	HD1	TRP	34	0.21	
VdW	O	LYS+	35 -	CB	CYSS	36	0.20	
VdW	O	LYS+	38 -	C	LYS+	39	0.31	
Hbond	HZ3	LYS+	2 -	O	CYSS	30	1	
Hbond	HN	GLN	3 -	O	TYR	1	1	
Hbond	HN	GLN	3 -	OD2	ASP	29	1	

```
Hbond HE21  GLN    3 - O     TYR    1      1
Hbond HE22  GLN    3 - N     GLY    8      1
Hbond HN    CYSS   4 - O     LYS+   2      1
Hbond HN    HIS+   5 - O     GLN    3      4
Hbond HD1   HIS+   5 - O     LEU   25      1
Hbond HN    LYS+   6 - O     CYSS   4      1
Hbond HN    LYS+   7 - O     HIS+   5      2
Hbond HZ3   LYS+   7 - O     CYSS   4      1
Hbond HN    GLY    8 - O     GLN    3      1
Hbond HN    GLY    8 - O     CYSS   4      1
Hbond HN    GLY    8 - O     HIS+   5      1
Hbond HN    GLY    9 - O     LYS+   6      1
Hbond HN    HIS+  10 - O     CYSS  37      6
Hbond HN    HIS+  10 - O     LYS+  38      2
Hbond HE2   HIS+  10 - O     LYS+  39      1
Hbond HN    CYSS  11 - OE1   GLN    3      1
Hbond HN    PHE   12 - OE1   GLU   15      1
Hbond HN    PHE   12 - O     LYS+  35      9
Hbond HN    LYS+  14 - O     PHE   12     15
Hbond HN    LYS+  14 - O     LYS+  33      1
Hbond HN    GLU   15 - O     PRO   13      1
Hbond HN    CYSS  18 - O     GLU   15      1
Hbond HN    CYSS  18 - OD2   ASP   24      1
Hbond HN    ILE   19 - O     LYS+  16      1
Hbond HN    ILE   19 - O     ILE   17      2
Hbond HN    SER   22 - O     PRO   20      5
Hbond HN    ASP   24 - O     SER   22      1
Hbond HN    ASP   24 - OD2   ASP   24      1
Hbond HD2   ASP   24 - O     PRO   20      1
Hbond HN    LEU   25 - O     CYSS  37      1
Hbond HN    GLY   26 - OD1   ASP   24      2
Hbond HN    GLY   26 - OD2   ASP   24      1
Hbond HN    LYS+  27 - N     CYSS  36      1
Hbond HZ1   LYS+  27 - OD1   ASP   24      1
Hbond HZ2   LYS+  27 - OD2   ASP   29      1
Hbond HZ2   LYS+  27 - O     ASP   29      1
Hbond HN    ASP   29 - O     LYS+  27      3
Hbond HD2   ASP   29 - N     TYR    1      1
Hbond HD2   ASP   29 - N     PHE   12      1
Hbond HN    CYSS  30 - O     MET   28      2
Hbond HN    CYSS  30 - O     TRP   34      1
Hbond HN    ARG+  31 - O     MET   28      1
Hbond HN    ARG+  31 - OD1   ASP   29      2
Hbond HN    TRP   32 - O     ASP   29      2
Hbond HN    LYS+  33 - OD1   ASP   29      1
Hbond HN    LYS+  33 - O     CYSS  30      1
Hbond HN    LYS+  33 - O     ARG+  31      4
Hbond HN    TRP   34 - O     ARG+  31      1
Hbond HN    LYS+  35 - O     PHE   12      8
Hbond HZ2   LYS+  35 - O     LYS+  33      1
Hbond HN    CYSS  36 - O     GLY   26      2
Hbond HN    CYSS  36 - N     LYS+  27      1
Hbond HN    CYSS  36 - O     LYS+  27      2
Hbond HN    CYSS  37 - N     HIS+  10      1
Hbond HN    CYSS  37 - O     HIS+  10      6
Hbond HN    LYS+  38 - O     SER   23      3
Hbond HN    LYS+  38 - OD1   ASP   24      1
Hbond HZ3   LYS+  38 - O     LYS+   7      1
Hbond HZ1   LYS+  39 - N     SER   23      1
Hbond HN    GLY   40 - O     LYS+  38      3
Hbond HN    SER   41 - O     ILE   19      1
Hbond HG    SER   41 - O     SER   41      1
Hbond HG    SER   41 - N     GLY   42      1
```

Pairwise RMSDs (structures ordered):

Number of backbone atoms : 102
Number of heavy atoms : 273
Residues considered : 6..39
Local RMSD segment length : 3 residues

Mean global backbone RMSD : 2.83 +/- 0.53 A (1.63..4.49 A)
Mean global heavy atom RMSD: 4.37 +/- 0.53 A (2.97..5.70 A)

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
1	484	2.52	2.30	2.41	2.86	2.82	2.51	2.39	3.20	2.74	3.56	2.76	2.57	3.30	3.45
2	4.10	653	2.30	2.39	2.31	3.15	2.78	2.69	2.25	2.25	3.30	1.63	2.53	2.26	2.75
3	3.79	4.20	796	2.46	1.94	2.09	1.81	2.61	2.26	2.23	3.30	2.05	2.62	2.75	3.20
4	3.87	4.29	4.16	902	2.81	3.56	2.51	1.89	2.83	2.55	2.48	2.46	3.08	3.60	3.06
5	4.42	4.14	3.65	4.52	342	2.82	1.99	2.73	1.91	2.22	3.29	2.13	3.10	3.06	2.74
6	4.16	4.51	3.65	5.17	4.19	755	2.94	3.20	3.28	2.74	4.12	2.96	2.98	2.95	3.85
7	4.11	4.63	3.41	4.09	3.44	4.21	663	2.96	2.26	2.67	3.32	2.59	3.17	3.52	3.36
8	4.12	4.16	3.94	3.89	4.27	4.47	4.61	479	3.21	2.90	2.64	2.79	2.70	3.45	2.77
9	4.76	3.77	4.10	4.50	4.20	4.93	3.81	4.88	766	2.33	3.33	2.22	3.33	3.14	3.16
10	4.10	3.77	4.30	4.11	4.02	4.22	3.97	4.74	3.55	157	3.39	2.13	3.13	3.01	3.38
11	4.76	4.87	4.57	3.86	4.91	5.59	4.49	4.08	4.63	4.45	562	3.06	3.92	3.81	2.70
12	4.61	2.97	3.97	4.24	4.10	4.26	4.46	4.20	3.65	3.54	4.58	365	3.05	2.60	2.79
13	3.91	3.92	3.94	4.47	4.57	4.54	4.75	3.68	4.84	4.77	5.01	4.69	240	2.39	3.48
14	4.35	3.24	4.39	5.08	4.64	4.20	5.14	4.22	4.64	4.51	5.05	4.06	3.45	313	3.24
15	5.30	4.56	4.88	5.05	4.56	5.45	5.27	4.38	5.08	5.15	4.36	4.72	4.46	4.27	836

Residue	local RMSD/segment		global RMSD/residue	
	backbone	heavy atoms	backbone	heavy atoms
1	0.00	0.00	10.93	12.42
2	0.95	2.69	8.73	10.34
3	0.93	2.29	6.35	7.71
4	0.89	2.33	3.54	3.41
5	0.89	2.30	3.84	6.32
6	0.87	2.74	4.39	6.99
7	0.84	2.20	4.08	5.95
8	0.74	1.88	3.80	3.99
9	0.74	1.56	2.10	2.18
10	0.62	1.44	1.45	3.07
11	0.40	1.49	1.28	1.49
12	0.27	0.97	1.42	2.29
13	0.52	1.78	1.91	2.25
14	0.69	2.00	2.22	4.79
15	0.71	2.35	2.90	5.41
16	0.80	2.15	2.38	4.55
17	0.57	1.74	2.55	3.94
18	0.74	1.63	1.61	1.85
19	0.45	1.04	3.20	4.22
20	0.14	0.70	3.45	3.82
21	0.54	0.98	2.47	3.10
22	0.59	1.18	2.21	2.78
23	0.61	1.43	2.46	3.10
24	0.69	1.67	2.12	3.18
25	0.72	1.66	3.29	5.18
26	0.84	2.04	2.66	2.81
27	0.79	2.18	2.60	5.49
28	0.89	2.55	3.03	4.84
29	0.85	2.17	3.10	4.99
30	0.77	2.16	1.80	2.02
31	0.72	2.31	2.22	4.29
32	0.56	2.51	2.83	4.99
33	0.43	2.26	1.95	3.06
34	0.33	1.75	1.13	2.26
35	0.28	1.41	1.01	2.61
36	0.41	1.39	1.05	1.41

37	0.64	1.69	1.19	1.65
38	0.85	2.22	2.34	4.18
39	0.90	2.28	4.85	6.96
40	0.81	1.93	7.72	8.17
41	0.87	1.32	10.57	11.31
42	0.00	0.00	12.95	13.43

APPENDIX D

GCG FASTA RESULTS

(Peptide) FASTA of: Myoa.Seq from: 1 to: 42 November 17, 1992 18:29

Myotoxin a from venom of Crotalus viridis viridis.

TO: SwissProt:* Sequences: 20,024 Symbols: 6,524,504 Word Size: 2

```
Score Init1 Initn
<  2   201   201:==================================================
   4    18    18:=========
   6    20    20:==========
   8    72    72:====================================
  10   544   544:==================================================
  12  1313  1313:==================================================
  14   200   200:==================================================
  16  1934  1934:==================================================
  18  3511  3511:==================================================
  20   559   559:==================================================
  22  4358  4358:==================================================
  24  4298  4298:==================================================
  26   447   447:==================================================
  28  1095  1078:==================================================
  30   562   554:==================================================
  32   199   191:==================================================
  34   249   232:==================================================
  36   160   160:==================================================
  38    83    88:=============================================++
  40   106   103:==================================================
  42    31    46:================+++++++
  44    27    37:===============+++++
  46    25    34:==============++++
  48     4     8:==++
  50     1     4:=+
  52     0     2:+
  54     0     2:+
  56     1     1:=
  58     0     1:+
  60     0     2:+
  62     0     0:
  64     0     0:
  66     0     0:
  68     0     0:
  70     0     0:
  72     0     0:
  74     0     0:
  76     0     0:
  78     0     0:
  80     0     0:
> 80     6     6:===
 mean initn score:  19.4 (4.89)
 mean init1 score:  19.4 (4.89)
```

The best scores are:		init1	initn	opt..
Sw:Myx1$Crovv	P01476 prairie rattlesnake (crotalus virid...	280	280	280
Sw:Myxc$Crodu	P01475 tropical rattlesnake (crotalus duri...	271	271	271
Sw:Myx2$Crovc	P12029 midget faded rattlesnake (crotalus ...	263	263	263
Sw:Myxc$Crovh	P01477 southern pacific rattlesnake (crota...	254	254	254
Sw:Myx1$Crovc	P12028 midget faded rattlesnake (crotalus ...	249	249	249
Sw:Myx2$Crovv	P19861 prairie rattlesnake (crotalus virid...	249	249	249
Sw:Bcr$Human	P11274 human (homo sapiens). breakpoint clu...	43	59	43
Sw:Li12$Caeel	P14585 caenorhabditis elegans. lin-12 prot...	48	59	53
Sw:Urok$Mouse	P06869 mouse (mus musculus). urokinase-typ...	40	58	41
Sw:Rnp$Macru	P00686 red kangaroo (macropus rufus). ribon...	55	55	55

```
Sw:Urok$Pig  P04185 pig (sus scrofa). urokinase-type plas...  36   54   39
Sw:Pol$Socmv  P15629 soybean chlorotic mottle virus. enzy...  39   53   39
Sw:Dpol$Bpt3  P20311 bacteriophage t3. dna polymerase (ec...  37   52   37
Sw:Nxl1$Denja  P01393 eastern jameson's mamba (dendroaspi...  42   51   42
Sw:Alb2$Xenla  P14872 african clawed frog (xenopus laevis...  36   50   36
Sw:Cin2$Rat  P04775 rat (rattus norvegicus). sodium chann...  50   50   51
Sw:Pol$Camvd  P03556 cauliflower mosaic virus (strain d/h...  42   49   42
Sw:Pol$Camvc  P03555 cauliflower mosaic virus (strain cm1...  42   49   42
Sw:Lmbs$Rat  P15800 rat (rattus norvegicus). s-laminin pr...  48   48   48
Sw:Vglm$Leev  P16853 lee virus. m polyprotein precursor (...  39   48   39
Sw:Vglm$Hojov  P16493 hojo virus. m polyprotein precursor...  39   48   39
Sw:Ompb$Chltr  P10553 chlamydia trachomatis. outer membra...  47   47   50
Sw:Itg1$Human  P20701 human (homo sapiens). leukocyte adh...  40   47   48
Sw:Nxl2$Denvi  P01395 western green mamba (dendroaspis vi...  38   47   38
Sw:Nxl1$Denvi  P01394 western green mamba (dendroaspis vi...  38   47   38
Sw:Cina$Eleel  P02719 electric eel (electrophorus electri...  47   47   47
Sw:Ibb2$Wheat  P09864 wheat (triticum aestivum). proteina...  46   46   51
Sw:Pa29$Pseau  P20253 mulga snake (pseudechis australis)....  34   46   41
Sw:Ibb2$Setit  P19860 foxtail millet (setaria italica). m...  46   46   47
Sw:Pa2c$Pseau  P20256 mulga snake (pseudechis australis)....  34   46   36
Sw:42$Human  P16452 human (homo sapiens). erythrocyte mem...  46   46   52
Sw:Pa2a$Pseau  P20255 mulga snake (pseudechis australis)....  34   46   36
Sw:Alk1$Human  P03973 human (homo sapiens). antileukoprot...  46   46   52
Sw:Gunb$Psefl  P18126 pseudomonas fluorescens. endoglucan...  46   46   48
Sw:Pa2b$Psepo  P20259 red-bellied black snake (pseudechis...  34   46   37
Sw:Ibb1$Wheat  P09863 wheat (triticum aestivum). proteina...  46   46   52
Sw:Pa2a$Psepo  P20258 red-bellied black snake (pseudechis...  34   46   36
Sw:Ibbr$Orysa  P07084 rice (oryza sativa). bran trypsin i...  46   46   48
Sw:Coat$Socmv  P15627 soybean chlorotic mottle virus. coa...  39   46   40
Sw:Pa21$Pseau  P04056 mulga snake (pseudechis australis)....  34   46   36


Myoa.Seq
Sw:Myx1$Crovv

ID   MYX1$CROVV     STANDARD;      PRT;    42 AA.
AC   P01476;
DT   21-JUL-1986  (REL. 01, CREATED)
DT   21-JUL-1986  (REL. 01, LAST SEQUENCE UPDATE)
DT   01-FEB-1991  (REL. 17, LAST ANNOTATION UPDATE)
DE   MYOTOXIN A (MYOTOXIN 1). . . .

SCORES      Init1: 280 Initn: 280 Opt: 280
            100.0% identity in 42 aa overlap

                10        20        30        40
Myoa.S YKQCHKKGGHCFPKEKICIPPSSDLGKMDCRWKWKCCKKGSG
       ||||||||||||||||||||||||||||||||||||||||||
Myx1$C YKQCHKKGGHCFPKEKICIPPSSDLGKMDCRWKWKCCKKGSG
                10        20        30        40


Myoa.Seq
Sw:Myxc$Crodu

ID   MYXC$CRODU     STANDARD;      PRT;    42 AA.
AC   P01475;
DT   21-JUL-1986  (REL. 01, CREATED)
DT   21-JUL-1986  (REL. 01, LAST SEQUENCE UPDATE)
DT   01-FEB-1991  (REL. 17, LAST ANNOTATION UPDATE)
DE   MYOTOXIN (CROTAMINE). . . .

SCORES      Init1: 271 Initn: 271 Opt: 271
             92.9% identity in 42 aa overlap
```

```
                10        20        30        40
Myoa.S YKQCHKKGGHCFPKEKICIPPSSDLGKMDCRWKWKCCKKGSG
       ||||||||||||||||||:|||||:||||||||:|||||||||
Myxc$C YKQCHKKGGHCFPKEKICLPPSSDFGKMDCRWRWKCCKKGSG
                10        20        30        40
```

Myoa.Seq
Sw:Myx2$Crovc

```
ID   MYX2$CROVC     STANDARD;      PRT;    43 AA.
AC   P12029;
DT   01-OCT-1989  (REL. 12, CREATED)
DT   01-OCT-1989  (REL. 12, LAST SEQUENCE UPDATE)
DT   01-FEB-1991  (REL. 17, LAST ANNOTATION UPDATE)
DE   MYOTOXIN II. . . .

SCORES     Init1: 263 Initn: 263 Opt: 263
            92.7% identity in 41 aa overlap

                10        20        30        40
Myoa.S YKQCHKKGGHCFPKEKICIPPSSDLGKMDCRWKWKCCKKGSG
       ||:||||||||||||||||:|||||:||||||||||||||||
Myx2$C YKRCHKKGGHCFPKEKICTPPSSDFGKMDCRWKWKCCKKGSVN
                10        20        30        40
```

Myoa.Seq
Sw:Myxc$Crovh

```
ID   MYXC$CROVH     STANDARD;      PRT;    43 AA.
AC   P01477;
DT   21-JUL-1986  (REL. 01, CREATED)
DT   21-JUL-1986  (REL. 01, LAST SEQUENCE UPDATE)
DT   01-FEB-1991  (REL. 17, LAST ANNOTATION UPDATE)
DE   MYOTOXIN (TOXIC PEPTIDE C). . . .

SCORES     Init1: 254 Initn: 254 Opt: 254
            87.8% identity in 41 aa overlap

                10        20        30        40
Myoa.S YKQCHKKGGHCFPKEKICIPPSSDLGKMDCRWKWKCCKKGSG
       ||:||||||||||||: ||:|||||:||||||||||||||||
Myxc$C YKRCHKKGGHCFPKTVICLPPSSDFGKMDCRWKWKCCKKGSVN
                10        20        30        40
```

Myoa.Seq
Sw:Myx1$Crovc

```
ID   MYX1$CROVC     STANDARD;      PRT;    43 AA.
AC   P12028;
DT   01-OCT-1989  (REL. 12, CREATED)
DT   01-OCT-1989  (REL. 12, LAST SEQUENCE UPDATE)
DT   01-FEB-1991  (REL. 17, LAST ANNOTATION UPDATE)
DE   MYOTOXIN I. . . .

SCORES     Init1: 249 Initn: 249 Opt: 249
            85.4% identity in 41 aa overlap

                10        20        30        40
Myoa.S YKQCHKKGGHCFPKEKICIPPSSDLGKMDCRWKWKCCKKGSG
       ||:||||:|||||||: ||:|||||:||||||||||||||||
Myx1$C YKRCHKKEGHCFPKTVICLPPSSDFGKMDCRWKWKCCKKGSVN
                10        20        30        40
```

```
Myoa.Seq
Sw:Myx2$Crovv

ID   MYX2$CROVV     STANDARD;      PRT;    45 AA.
AC   P19861;
DT   01-FEB-1991  (REL. 17, CREATED)
DT   01-FEB-1991  (REL. 17, LAST SEQUENCE UPDATE)
DT   01-FEB-1991  (REL. 17, LAST ANNOTATION UPDATE)
DE   MYOTOXINS 2 AND 3. . . .

SCORES     Init1: 249 Initn: 249 Opt: 249
            85.4% identity in 41 aa overlap

                10        20        30        40
Myoa.S YKQCHKKGGHCFPKEKICIPPSSDLGKMDCRWKWKCCKKGSG
       ||:||||:||||||: ||:|||||:||||||||||||||||
Myx2$C YKRCHKKEGHCFPKTVICLPPSSDFGKMDCRWKWKCCKKGSVNNA
                10        20        30        40


Myoa.Seq
Sw:Bcr$Human

ID   BCR$HUMAN      STANDARD;      PRT;  1271 AA.
AC   P11274;
DT   01-JUL-1989  (REL. 11, CREATED)
DT   01-JUL-1989  (REL. 11, LAST SEQUENCE UPDATE)
DT   01-FEB-1991  (REL. 17, LAST ANNOTATION UPDATE)
DE   BREAKPOINT CLUSTER REGION PROTEIN (GENE NAME: BCR). . . .

SCORES     Init1: 43 Initn: 59 Opt: 43
            66.7% identity in 9 aa overlap

                                          10        20        30
Myoa.S                                YKQCHKKGGHCFPKEKICIPPSSDLGKMDCRW
                                        ||||:  ||
Bcr$Hu PTTYRMFRDKSRSPSQNSQQSFDSSSPPTPQCHKRHRHCPVVVSEATIVGVRKTGQIWPN
       360       370       380       390       400       410

            40
Myoa.S KWKCCKKGSG

Bcr$Hu DGEGAFHGDADGSFGTPPGYGCAADRAEEQRRHQDGLPYIDDSPSSSPHLSSKGRGSRDA
       420       430       440       450       460       470


Myoa.Seq
Sw:Li12$Caeel

ID   LI12$CAEEL     STANDARD;      PRT;  1429 AA.
AC   P14585;
DT   01-JAN-1990  (REL. 13, CREATED)
DT   01-JAN-1990  (REL. 13, LAST SEQUENCE UPDATE)
DT   01-FEB-1991  (REL. 17, LAST ANNOTATION UPDATE)
DE   LIN-12 PROTEIN PRECURSOR (GENE NAME: LIN12). . . .

SCORES     Init1: 48 Initn: 59 Opt: 53
            30.8% identity in 26 aa overlap

                                       10        20        30
Myoa.S                              YKQCHKKGGHCFPKEKICIPPSSDLGKMDCRWKW
                                      :|:::   :::| ||::::|:: |||
Li12$C TEPITRESVNIIDPRHNRTVLHWIASNSSAEKSEDLIVHEAKECIAAGADVNAMDCDENT
```

```
       1040      1050      1060      1070      1080      1090

          40
Myoa.S KCCKKGSG

Li12$C PLMLAVLARRRRLVAYLMKAGADPTIYNKSERSALHQAAANRDFGMMVYMLNSTKLKGDI
       1100      1110      1120      1130      1140      1150


Myoa.Seq
Sw:Urok$Mouse

ID   UROK$MOUSE     STANDARD;      PRT;   433 AA.
AC   P06869;
DT   01-JAN-1988  (REL. 06, CREATED)
DT   01-JAN-1988  (REL. 06, LAST SEQUENCE UPDATE)
DT   01-APR-1990  (REL. 14, LAST ANNOTATION UPDATE)
DE   UROKINASE-TYPE PLASMINOGEN ACTIVATOR PRECURSOR (EC 3.4.21.31) (UPA). . . .

SCORES     Init1: 40 Initn: 58 Opt: 41
            44.4% identity in 18 aa overlap

                    10        20        30        40
Myoa.S       YKQCHKKGGHCFPKEKICIPPSSDLGKMDCRWKWKCCKKGSG
                                       |   |  || ::|| :||
Urok$M KNLKMSVVKLVSHEQCMQPHYYGSEINYKMLCAADPEWKTDSCKGDSGGPLICNIEGRPT
            340       350       360       370       380       390

Urok$M LSGIVSWGRGCAEKNKPGVYTRVSHFLDWIQSHIGEEKGLAF
            400       410       420       430


Myoa.Seq
Sw:Rnp$Macru

ID   RNP$MACRU      STANDARD;      PRT;   122 AA.
AC   P00686;
DT   21-JUL-1986  (REL. 01, CREATED)
DT   21-JUL-1986  (REL. 01, LAST SEQUENCE UPDATE)
DT   01-MAR-1989  (REL. 10, LAST ANNOTATION UPDATE)
DE   RIBONUCLEASE PANCREATIC (EC 3.1.27.5) (RNASE 1) (RNASE A). . . .

SCORES     Init1: 55 Initn: 55 Opt: 55
            28.6% identity in 28 aa overlap

                                        10        20        30
Myoa.S                               YKQCHKKGGHCFPKEKICIPPSSDLGKMDCRWK
                                       ||:::  |    : |  ::| |:  :||
Rnp$Ma LMMKARDMTSGRCKPLNTFIHEPKSVVDAVCHQENVTCKNGRTNCYKSNSRLSITNCRQT
         30        40        50        60        70        80

          40
Myoa.S WKCCKKGSG

Rnp$Ma GASKYPNCQYETSNLNKQIIVACEGQYVPVHFDAYV
         90       100       110       120


Myoa.Seq
Sw:Urok$Pig

ID   UROK$PIG       STANDARD;      PRT;   442 AA.
AC   P04185;
DT   20-MAR-1987  (REL. 04, CREATED)
DT   13-AUG-1987  (REL. 05, LAST SEQUENCE UPDATE)
```

```
DT   01-APR-1990  (REL. 14, LAST ANNOTATION UPDATE)
DE   UROKINASE-TYPE PLASMINOGEN ACTIVATOR PRECURSOR (EC 3.4.21.31) (UPA). . . .

SCORES     Init1: 36 Initn: 54 Opt: 39
            38.9% identity in 18 aa overlap

                      10        20        30        40
Myoa.S       YKQCHKKGGHCFPKEKICIPPSSDLGKMDCRWKWKCCKKGSG
                                      |   | :|| ::|: :||
Urok$P EQLKMTVVKLVSHRECQQPHYYGSEVTTKMLCAADPQWKTDSCQGDSGGPLVCSTQGRLT
              350       360       370       380       390       400

Urok$P LTGIVSWGRECAMKDKPGVYTRVSRFLTWIHTHVGGENGLAH
              410       420       430       440


Myoa.Seq
Sw:Pol$Socmv

ID   POL$SOCMV      STANDARD;      PRT;   741 AA.
AC   P15629;
DT   01-APR-1990  (REL. 14, CREATED)
DT   01-APR-1990  (REL. 14, LAST SEQUENCE UPDATE)
DT   01-APR-1990  (REL. 14, LAST ANNOTATION UPDATE)
DE   ENZYMATIC POLYPROTEIN (CONTAINS: ASPARTIC PROTEASE (EC 3.4.23.-), . . .

SCORES     Init1: 39 Initn: 53 Opt: 39
           100.0% identity in 3 aa overlap

             10        20        30        40
Myoa.S KQCHKKGGHCFPKEKICIPPSSDLGKMDCRWKWKCCKKGSG
                                     |||
Pol$So CINYIAPEGFFRTLALERKHLQKKISVKNPWKWDTIDTKMVQSIKGKIQSLPKLYNASIQ
         450       460       470       480       490       500

Pol$So DFLIVETDASQHSWSGCLRALPRESKKSDSMNSGYRPCDLCTGSSSASSDNSPAEIDKCH
         510       520       530       540       550       560


Myoa.Seq
Sw:Dpol$Bpt3

ID   DPOL$BPT3      STANDARD;      PRT;   704 AA.
AC   P20311;
DT   01-FEB-1991  (REL. 17, CREATED)
DT   01-FEB-1991  (REL. 17, LAST SEQUENCE UPDATE)
DT   01-FEB-1991  (REL. 17, LAST ANNOTATION UPDATE)
DE   DNA POLYMERASE (EC 2.7.7.7) (GENE NAME: 5). . . .

SCORES     Init1: 37 Initn: 52 Opt: 37
            62.5% identity in 8 aa overlap

                      10        20        30        40
Myoa.S       YKQCHKKGGHCFPKEKICIPPSSDLGKMDCRWKWKCCKKGSG
                                     |||: :||
Dpol$B EIAKTVIEVAQEAMRWVGEHWNFRCLLDTEGKMGANWKECH
           670       680       690       700


Myoa.Seq
Sw:Nxl1$Denja

ID   NXL1$DENJA     STANDARD;      PRT;    72 AA.
AC   P01393;
DT   21-JUL-1986  (REL. 01, CREATED)
```

```
DT   21-JUL-1986  (REL. 01, LAST SEQUENCE UPDATE)
DT   01-APR-1988  (REL. 07, LAST ANNOTATION UPDATE)
DE   LONG NEUROTOXIN 1. . . .

SCORES      Init1: 42 Initn: 51 Opt: 42
             38.9% identity in 18 aa overlap

                    10        20        30        40
Myoa.S     YKQCHKKGGHCFPKEKICIPPSSDLGKMDCRWKWKCCKKGSG
           || : :|:  | : |:||
Nx11$D RTCYKTYSDKSKTCPRGEDICYTKTWCDGFCSQRGKRVELGCAATCPKVKTGVEIKCCST
                10        20        30        40        50        60


Myoa.Seq
Sw:Alb2$Xenla

ID   ALB2$XENLA      STANDARD;       PRT;   607 AA.
AC   P14872;
DT   01-APR-1990  (REL. 14, CREATED)
DT   01-NOV-1990  (REL. 16, LAST SEQUENCE UPDATE)
DT   01-NOV-1990  (REL. 16, LAST ANNOTATION UPDATE)
DE   74 KD SERUM ALBUMIN PRECURSOR. . . .

SCORES      Init1: 36 Initn: 50 Opt: 36
             60.0% identity in 10 aa overlap

                                              10        20        30
Myoa.S                                  YKQCHKKGGHCFPKEKICIPPSSDLGKMDCR
                                        ||:|| ::||
Alb2$X EHPDDLLSAFIHEEARNHPDLYPPAVLALTKQYHKLAEHCCEEEDKEKCFSEKMKQLMKQ
             160       170       180       190       200       210

              40
Myoa.S WKWKCCKKGSG

Alb2$X SHSIEDKQHHFCWILDNFPEKVLKALNLARVSHRYPKAEFKLAHNFTEEVTHFIKDCCHD
             220       230       240       250       260       270


Myoa.Seq
Sw:Cin2$Rat

ID   CIN2$RAT        STANDARD;       PRT;  2005 AA.
AC   P04775;
DT   13-AUG-1987  (REL. 05, CREATED)
DT   13-AUG-1987  (REL. 05, LAST SEQUENCE UPDATE)
DT   01-APR-1990  (REL. 14, LAST ANNOTATION UPDATE)
DE   SODIUM CHANNEL PROTEIN II, BRAIN. . . .

SCORES      Init1: 50 Initn: 50 Opt: 51
             60.0% identity in 10 aa overlap

                  10        20        30        40
Myoa.S    YKQCHKKGGHCFPKEKICIPPSSDLGKMDCRWKWKCCKKGSG
                                        || :|:|||:
Cin2$R STVDIGAPAEGEQPEAEPEESLEPEACFTEDCVRKFKCCQ:SIEEGKGKLWWNLRKTCYK
          1150      1160      1170      1180      1190      1200

Cin2$R IVEHNWFETFIVFMILLSSGALAFEDIYIEQRKTIKTMLEYADKVFTYIFILEMLLKWVA
          1210      1220      1230      1240      1250      1260
```